



GMD Research Series

GMD –
Forschungszentrum
Informationstechnik
GmbH

Oliver Eulenstein

Vorhersage von Genduplikationen und deren Entwicklung in der Evolution

© GMD 1998

GMD – Forschungszentrum Informationstechnik GmbH

Schloß Birlinghoven

D-53754 Sankt Augustin

Germany

Telefon +49 -2241 -14 -0

Telefax +49 -2241 -14 -2618

<http://www.gmd.de>

In der Reihe GMD Research Series werden Forschungs- und
Ergebnisse aus der GMD zum wissenschaftlichen, nicht-
kommerziellen Gebrauch veröffentlicht. Jegliche Inhaltsänderung des
Dokuments sowie die entgeltliche Weitergabe sind verboten.

The purpose of the GMD Research Series is the dissemination
of research work for scientific non-commercial use.

The commercial distribution of this document is prohibited,
as is any modification of its content.

Anschrift des Verfassers/Address of the author:

Oliver Eulenstein

Rheinische Friedrich-Wilhelms-Universität Bonn

Römerstraße 164

53 Bonn

E-mail: oliver.eulenstein@gmd.de

Die vorliegende Veröffentlichung entstand im/

The present publication was prepared within:

Institut für Algorithmen und Wissenschaftliches Rechnen (SCAI)

Institute for Algorithms and Scientific Computing

<http://www.scai.gmd.de>

D5

Die Deutsche Bibliothek - CIP-Kurztitelaufnahme:

Eulenstein, Oliver:

Vorhersage von Genduplikationen und deren Entwicklung in der
Evolution / Oliver Eulenstein. GMD – Forschungszentrum
Informationstechnik GmbH. - Sankt Augustin :

GMD – Forschungszentrum Informationstechnik, 1998

(GMD Research Series ; 1998, No.20)

Zugl.: Bonn, Univ., Diss., 1998

ISBN 3-88457-344-6

ISSN 1435-2699

ISBN 3-88457-344-6

Abstract

Evolution proceeds via gene duplication and modification; it is through their successive application that nature has created the vast diversity of current genes. When a gene encoding an essential function is duplicated, one copy must preserve the function for its organism. The other, however, is free to be modified, using its ancestral gene as its starting point. Knowing precisely when ancestral gene duplications occurred is indispensable for reconstructing reliable phylogenetic relationships which in turn allows us to refine the prediction of a genetic function. For some gene families, such as globins or rhodopsins, ancestral gene duplications are well predicted, but in general they are unknown. Mapping out this unknown territory is an important open issue in molecular biology today.

Recognizing this, we designed an algorithm that extends a successful but informal model built by biologists. This algorithm predicts gene duplications by exploiting inconsistencies in two phylogenetic trees, a gene tree and a species tree. The gene tree represents the phylogeny of a set of current genes, while the species tree represents the phylogeny of the current genes organized by their host-species. As output the algorithm calculates the gene tree reconciled with its predicted gene duplications, which is called a reconciled tree. For all applications that are possible in practice, this algorithm runs in linear time in the size of the given gene tree. This time complexity allows us to screen the today's fast growing molecular databases for gene duplications. First, we partition a genetic database into gene-families and reconstruct a gene tree for each family by one of the well established and fast reconstruction algorithms. For each gene tree, we then find its species tree in a morphologic database. Finally we predict the gene duplications for every gene and species tree pair. Thus, we are able to predict hidden gene duplications throughout the genetic database.

While our algorithm predicts gene duplications, it simultaneously associates a reliability index with the set of predictions it has output. Since the reconciled tree is quadratic in the size of the gene tree, we developed a new graphical representation to ease its analysis by biologists. In essence, this graphical representation is the gene tree reconciled only at those gene duplications that are of interest for the biologist. Through this logical compression in presentation, we can, in general, avoid the quadratic explosion. Up to now, this work is of theoretical nature, but its transformation into applications for biologists has begun.

Keywords: Phylogenetic tree; Gene-tree; Species-tree; Reconciled-tree; Measures for trees; Gene duplication

Evolution vollzieht sich durch Genduplikation und Modifikation, durch deren wiederholte Anwendung die Natur die Vielzahl der heutigen unterschiedlichen Gene geschaffen hat. Wird ein Gen das eine wesentliche Funktion kodiert, dupliziert, so bewahrt eine Kopie diese Funktion für dessen Organismus. Die andere Kopie jedoch kann ungehindert modifiziert werden, wobei das Ursprungs-Gen dazu als Ausgangsbasis dient. Eine genaue Kenntnis vorausgegangener Genduplikationen ist für die zuverlässige Rekonstruktion phylogenetischer Beziehungen unerlässlich, welche uns dann erlaubt, die Funktionsvorhersage von Genen zu verbessern. Für einige Genfamilien, wie z.B. Globine oder Rhodopsine, sind Genduplikationen glaubhaft vorhergesagt worden, aber im allgemeinen sind diese noch unbekannt. Die Erforschung dieses unbekanntes Gebietes stellt ein bedeutendes, noch ungelöstes Problem der heutigen Molekularbiologie dar. Um dieses Problem zu lösen, konstruierten wir einen Algorithmus, der ein erfolgreiches aber nicht formal beschriebenes Modell erweitert, welches ursprünglich von Biologen entwickelt wurde. Dieser Algorithmus sagt Genduplikationen durch die Auswertung der Inkonsistenzen zwischen zwei phylogenetischen Bäumen voraus, einem Genbaum und einem Speziesbaum. Der Genbaum stellt die Phylogenie für eine Menge aktueller Gene dar, während der Speziesbaum die Phylogenie der Wirts-Spezies dieser aktuellen Gene darstellt. Als Ausgabe berechnet der Algorithmus einen auf dessen vorhergesagte Genduplikationen abgestimmten Genbaum, welchen wir als *Reconciled Tree* bezeichnen. Für alle in der Praxis möglichen Anwendungen hat dieser Algorithmus ein lineares Laufzeitverhalten, in der Größe des gegebenen Genbaumes. Diese Zeitkomplexität ermöglicht es uns, die heutigen, schnell wachsenden Gendatenbanken nach Genduplikationen zu durchsuchen. Zuerst partitionieren wir dazu eine Gendatenbank in Genfamilien und rekonstruieren einen Genbaum für jede dieser Genfamilien durch einen etablierten und schnellen Rekonstruktions-Algorithmus. Dann suchen wir für jeden Genbaum seinen Speziesbaum in einer morphologischen Datenbank. Abschließend sagen wir die Genduplikationen für jedes Paar von Gen- und Speziesbaum voraus. Somit können wir in der Gendatenbank verborgene Genduplikationen vorhersagen.

Während unser Algorithmus Genduplikationen vorhersagt, ordnet er gleichzeitig der Menge seiner ausgegebenen Vorhersagen einen Zuverlässigkeits-Index zu. Da der *Reconciled Tree* quadratisch in der Größe des Genbaumes ist, entwickelten wir eine neue graphische Darstellung, um die Analyse durch den Biologen zu vereinfachen. Im Wesentlichen ist diese graphische Darstellung der Genbaum, welcher nur auf die für den Biologen interessanten Genduplikationen abgestimmt ist. Durch diese natürliche Reduktion der Darstellung können wir im allgemeinen die quadratische Explosion vermeiden. Bis jetzt ist diese Arbeit theoretischer Natur, die Umwandlung in die Anwendung hat jedoch begonnen.

Schlagerworte: Phylogenetischer Baum; Genbaum; Speziesbaum; Reconciled-Tree; Maße für Bäume; Genduplikation

Vorwort

Diese Arbeit ist während meiner Tätigkeit als wissenschaftlicher Mitarbeiter an der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn zwischen 1994 und 1998 entstanden.

Mein besonderer Dank gilt meinem Doktorvater Prof. Dr. Thomas Lengauer, Ph.D. Er begeisterte mich für das Gebiet der molekularen Bioinformatik, ermöglichte mir das wissenschaftliche Arbeiten auf diesem Gebiet, war Ansprechpartner und hat meine Arbeit kritisch begutachtet.

Im besonderen verdanke ich ihm einen großen Forschungsfreiraum, der eine wesentliche Grundlage für mein wissenschaftliches Arbeiten ist. Von großer Bedeutung für mich ist aber auch die Zusammenarbeit mit Kollegen der Forschungsgruppe von Prof. Dr. Thomas Lengauer, welche einen wichtigen Einfluß auf meine Arbeit hatten.

Initiiert wurde diese Arbeit im Sommer 1994 durch Dr. Martin Vingron, worauf eine enge Zusammenarbeit mit ihm folgte. Nicht zuletzt diese Zusammenarbeit ermöglichte mir die vorliegende Arbeit zu erstellen. Ich profitierte in wesentlichen Teilen meiner Arbeit von seinem Wissen und seinen Erfahrungen. Intensive Diskussionen mit ihm ermöglichten mir oft einen neuen Zugang zu einem Problem oder brachten Klarheit in die Darstellung meiner Resultate.

Die ersten Resultate dieser Arbeit weckten das Interesse von Dr. Frieder Kaden. Sein tiefes mathematisches Verständnis half mir, Fehler in meinen Schlußfolgerungen aufzudecken und mir selbst klarer über die Problematik zu werden.

Nicht vergessen möchte ich Benno Schwikowski, mit welchem ich zahlreiche und unermüdliche Diskussionen zu Problemen wie auch Resultaten meiner Arbeit führte. Diese gaben mir insbesondere dann neue Anregungen, wenn ich dachte ein Problem nicht mehr lösen zu können, aber im Gespräch dann doch neue Lösungswege sah.

Abschließend danke ich Prof. Dr. Wolfgang Alt für seine Mühe, die er durch meine Betreuung auf sich genommen hat.

Inhaltsübersicht

1	Einleitung	1
1.1	Die Bedeutung der Genduplikation	1
1.1.1	Zuordnung von Funktionalität	2
1.1.2	Rekonstruktion von Genbäumen	4
1.2	Die Vorhersage von Genduplikationen	6
1.2.1	Ein Reconciled Tree nach Page	6
1.2.2	Strukturaussagen zu einem Reconciled Tree	8
1.2.3	Der konstruktive Reconciled Tree	9
1.2.4	Maße zum konstruktiven Reconciled Tree	10
1.2.5	Algorithmus	11
1.2.6	Anwendung	11
2	Darstellung von phylogenetischen Bäumen	13
2.1	Die Darstellung als Mengensystem	14
2.2	Das n-Clustersystem	16
2.2.1	Die Baumordnung	16
2.2.2	Das n-Clustersystem	17
2.3	Die Erweiterung der kleinsten oberen Schranke	18
3	Der Reconciled Tree nach R. Page	21
3.1	Die Ableitung aus der Evolution	21
3.2	Die historische Definition	24
3.2.1	Ableitung, als n-Clustersystem einer Multimenge	24
3.2.2	Das m-Clustersystem	26
3.2.3	Der Reconciled Tree	27
3.3	Eine formale Definition	27
3.3.1	Reformulierung der Multimenge	28
3.3.2	Reformulierung des m-Clustersystemes	30
3.3.3	Reformulierung des Reconciled Tree	31

4	Strukturaussagen zum Reconciled Tree	35
4.1	Operationen auf m-Clusternsystemen	36
4.1.1	Die Operation: Replace	36
4.1.1.1	Formulierung der Ersetzung	37
4.1.1.2	Wohldefiniertheit der Ersetzung	39
4.1.2	Die Operation: Merge	45
4.1.3	Die Operation: \times	47
4.2	Die Ableitung von Genbäumen	47
4.3	Das LCA-Theorem	50
4.3.1	Beweisskizze	51
4.3.1.1	Struktureller Abschnitt	51
4.3.1.2	Konstruktiver Abschnitt	52
4.3.2	Der formale LCA-Beweis	56
4.4	Die Topologie des Reconciled Tree	65
4.4.1	Die Entwicklungsumgebungen von \mathcal{G}	65
4.4.1.1	Der topologische Bezug von $E_{\mathcal{R}}(X)$ zu \mathcal{S}	67
4.4.1.2	Duplikationselemente in $E_{\mathcal{R}}(X)$	68
4.4.2	Regionen einer Entwicklungsumgebung	69
4.4.2.1	Der Bezug der Regionen von $E_{\mathcal{R}}(X)$ zu $E_{\mathcal{S}}(X)$	71
4.4.3	Die Topologie einer Entwicklungsumgebung $E_{\mathcal{R}}(X)$	74
4.4.3.1	Die Topologie von $E_{\mathcal{R}}(X)$, wenn $P_{\text{NoDup}}(X)$ gilt	74
4.4.3.2	Das Duplikations-Theorem	79
4.4.3.3	Die Topologie von $E_{\mathcal{R}}(X)$, wenn $\neg P_{\text{NoDup}}(X)$ gilt	82
5	Konstruktive Definitionen zum Reconciled Tree	89
5.1	Duplikationselemente in \mathcal{G}	89
5.2	Der Reconciled Tree, konstruktiv definiert	90
5.2.1	Der konstruktive Reconciled Tree: Die Idee	91
5.2.1.1	Zuordnung des ersten Dreitupelparameters	92
5.2.1.2	Zuordnung des zweiten Dreitupelparameters	93
5.2.1.3	Zuordnung des dritten Dreitupelparameters	95
5.2.2	Die konstruktive Reconciled Tree: formal	96
5.2.2.1	Formale Definition der Dreitupel-Bezeichner	96
5.2.2.2	Die Formale Definition einer Ordnung auf RK	99
5.3	Ausblick: Der Mixed Tree	104
6	Die biologischen Maße des Reconciled Tree	107
6.1	Maße für UGs	108
6.1.1	Ein einfaches Maß für UGs	110

6.1.2	Maße realer UGs	110
6.1.2.1	Das Maß M_1	110
6.1.2.2	Das Maß M_f	113
6.2	Die Historie der Maße zum Reconciled Tree	118
6.3	Die Mirkin-Muchnik-Smith-Vermutung	120
6.3.1	Das Maß M_{MMS}	121
6.3.2	Ein Beispiel zu der MMS-Vermutung	124
6.3.2.1	Die Berechnung des Maßes $M_{MMS}(\mathcal{G}, \mathcal{S})$	125
6.3.2.2	Die Berechnung des Maßes $M_{Muchnik}(\mathcal{G}, \mathcal{S})$	126
6.3.3	Eine Idee zum Beweis der MMS-Vermutung	126
6.4	Ausblick	128
7	Algorithmen zur Vorhersage von Genduplikationen	131
7.1	Eine Lösung des Dup-Problemes	132
7.2	Eine direkte Lösung des LCA-Problemes	134
7.2.1	Der LCA-Algorithmus	134
7.2.1.1	Die Prozedur SPECIES TREE-DFS	135
7.2.1.2	Die Prozedur GENE TREE-WALK	136
7.2.2	Eine Verifikation des LCA-Algorithmus	137
7.2.3	Zeitkomplexität	141
7.3	Eine Reduktion des LCA-Problemes	143
7.3.1	Eine Linearzeit-Reduktion	143
7.3.2	Berechnungen von \sqcup	143
7.4	Die Berechnung des Maßes M_1	145
8	Paraloge Evolution in der Farbwahrnehmung	147
8.1	Konstruktion des Reconciled Tree	149
8.1.1	Homologe Sequenzen zu OPSB_ANOCA	149
8.1.2	Rekonstruktion von \mathcal{G} und \mathcal{S}	151
8.1.3	Bestimmung des Reconciled Tree für \mathcal{G} und \mathcal{S}	154
8.1.3.1	Notwendige Definitionen für \mathcal{RK}	155
8.1.3.2	Bestimmung der Menge \mathcal{RK}	156
8.1.3.3	Hasse-Diagramm der Baumordnung \subseteq auf \mathcal{RK}	158
8.2	Analyse des Reconciled Tree	159
9	Schluß: Zusammenfassung und Ausblick	163
9.1	Zusammenfassung	163
9.2	Ausblick	165
	Allgemeine Definitionen	167

Abbildungen

1.1	Paraloge Gene.	3
1.2	Vollständiger Genbaum	4
1.3	Inkonsistenz: korrekter Gen- und Speziesbaum	4
1.4	Entwicklung der Kopien im Speziesbaum	5
1.5	Die Multifurkation zeigt Unsicherheiten	8
2.1	Die Hasse-Diagramme der Mengensysteme von \mathcal{G} und \mathcal{S}	13
2.2	Ein Genbaum \mathcal{G} und Speziesbaum \mathcal{S}	19
3.1	Inkonsistenz eines korrekten Genbaumes \mathcal{G} und Speziesbaumes \mathcal{S}	22
3.2	Die Ableitung \mathcal{R} eines zu \mathcal{S} konsistenten Genbaumes	23
3.3	Eine spezieskonforme Einbettung von \mathcal{G} in \mathcal{R}	24
3.4	Die Ableitung des Genbaumes \mathcal{G}_j durch das Duplikationselement α aus dem Speziesbaum \mathcal{S}	25
3.5	Der formal definierte Reconciled Tree \mathcal{R} von \mathcal{G} über \mathcal{S}	28
4.1	Die Ersetzung $\text{Repl}(\mathcal{M}, \mathcal{N}, I)$	37
4.2	Die Partitionierung von \mathcal{M} und der Ersetzung $\mathcal{L} := \text{Repl}(\mathcal{M}, \mathcal{N}, I)$	40
4.3	Die Zusammenfügung $\text{Merge}(\mathcal{N}_1, \dots, \mathcal{N}_n)$	45
4.4	Strukturelle Eigenschaften	51
4.5	Konstruktion von \mathcal{R}''	52
4.6	Der Einbettungsmonomorphismus φ_{neu} von \mathcal{G} nach \mathcal{R}''	54
4.7	Zu den Kardinalitäten $ \mathcal{R} $ und $ \mathcal{R}'' $	55
4.8	Die Entwicklungsumgebungen von \mathcal{R}	66
4.9	Die lokalen Entwicklungsumgebungen von $E_{\mathcal{R}}(\mathbf{b})$	70
4.10	Die Konstruktion von \mathcal{R}^n	75
4.11	Die Konstruktion von \mathcal{R}''	84
5.1	Der konstruktive Reconciled Tree $\text{RK}_{\mathcal{G}, \mathcal{S}}$	91
5.2	Der konstruktive Reconciled Tree $\text{RK}(\mathcal{G}, \mathcal{S})$	92
5.3	Die Dreitupelbezeichnungen von $E_{\mathcal{R}'}(\mathbf{b})$	93
5.4	Die Dreitupelbezeichnungen von $\tilde{E}_{\mathcal{R}}(e)$	94

5.5	Fall 1a links und Fall 1b rechts	101
5.6	Fall 2a links und Fall 2b rechts	102
5.7	Ein Reconciled Tree und der konstruktive Reconciled Tree	103
5.8	Ein Mixed Tree	104
6.1	BGs und UGs in einem konstruktiven Reconciled Tree	109
6.2	MUTs in einem konstruktiven Reconciled Tree	113
6.3	MUT in der unbestimmten Entwicklungsumgebung $\tilde{E}_{\mathcal{R}}(e)$	114
6.4	Die lokale Entwicklung der Duplikate von a	122
6.5	Genbaum und Speziesbaum	124
6.6	Die Berechnung des Maßes $M_{\text{MMS}}(\mathcal{G}, \mathcal{S})$	125
8.1	Multiples Alignment	152
8.2	Ungewurzelter Genbaum	153
8.3	Gewurzelter Genbaum \mathcal{G}	153
8.4	Speziesbaum \mathcal{S}	154
8.5	Zur Konstruktion von UG_1 (dabei gilt $y' = \sqcup_{\mathcal{S}}(y)$)	157
8.6	Zur Konstruktion von UG_2 (dabei gilt $y' = \sqcup_{\mathcal{S}}(y)$)	158
8.7	Der Reconciled Tree \mathcal{RK}	159
8.8	Eine Hypothese am Reconciled Tree \mathcal{RK}	160
9.1	Die Ersetzung $\text{Repl}(\mathcal{M}, \mathcal{N}, I)$	172
9.2	Konstruktion der Abbildung φ''	174

Kapitel 1

Einleitung

Die Evolution entwickelt Funktionalität allgemein durch die Duplikation von Genen (Nei, 1987 Seite 409; Li und Graur, 1991). Bei diesem einfachen aber effizienten Mechanismus sichert sich die Evolution durch eine Kopie einer Genduplikation die vorhandene Funktionalität und verwendet die andere Kopie, um neue Funktionalität zu erschaffen. Viele der uns heute bekannten Gene gehören zu großen und inzwischen intensiv studierten Genfamilien, wie sie sich aus einer Reihe von Genduplikationen entwickelt haben. Damit ist die Genduplikation ein fundamentaler Mechanismus der genomischen Evolution und eine wesentliche Voraussetzung für die funktionale Diversifizierung.

Wir werden in dieser Arbeit ein direkt aus der Evolution abgeleitetes Modell verwenden, für das wir einen schnellen Algorithmus konstruieren, um damit Genduplikationen und deren Entwicklung in der Evolution vorherzusagen, sowie eine Qualität dieser Vorhersage angeben.

In dem folgenden Abschnitt 1.1 zeigen wir die Auswirkungen der Genduplikation und damit die Bedeutung ihrer Vorhersage für die Molekularbiologie und Phylogenie. Einen einführenden Überblick über die Vorhersage von Genduplikationen sowie die Struktur dieser Arbeit geben wir in Abschnitt 1.2. Eine Einführung in die übliche Darstellung phylogenetischer Bäume ist in Kapitel 2 gegeben. Allgemein bekannte Begriffe der Mathematik, die wir in dieser Arbeit verwenden, sind in Anhang 9.2 zu finden.

1.1 Die Bedeutung der Genduplikation

Vereinfacht sind Gene eines gemeinsamen Vorgängergens paralog, wenn sie aus einer Duplikation dieses Vorgängergens hervorgegangen sind, ortholog. Auf Grund der funktionalen Implikationen, welche das Wissen um Gruppen ortho-

loger Gene mit sich bringt, besteht ein großes Interesse an deren Erkennung (z.B. Fitch, 1970; Rosen, 1978; Goodman et al., 1979; Nelson und Platnick, 1981; Page, 1990b; Page, 1993a; Page, 1994a; Register et al., 1994; Kawamura und Yokohama, 1995; Guigo et al., 1996; Tatusov et al., 1997; Slowinski et al., 1997; Yuan et al., 1998).

Neben der Erkennung der orthologen Gene ist aber auch das Wissen um paraloge Gene von großer Bedeutung für die experimentelle Funktionsbestimmung. So ist der Erfolg von *knock-out* Experimenten u.a. davon abhängig, ob die Funktion eines ausgeschalteten Gens durch ein ähnliches Gen, möglicherweise eine andere Kopie, übernommen werden kann. In diesem Falle könnte ein sehr aufwendiges Experiment ohne Ergebnis bleiben.

In den folgenden Abschnitten werden wir zuerst diese Begriffe und deren Relevanz genauer diskutieren. Dann geben wir eine Einführung in die Darstellung von phylogenetischen Bäumen zum Zwecke der Bestimmung von Genduplikationen.

1.1.1 Zuordnung von Funktionalität

Rapide Fortschritte der weltweiten Sequenzierung von Genomen stellen eine massiv anwachsende Menge an Gensequenzen zur Verfügung. Aber erst deren präzise Annotation ermöglicht es, diese Information sinnvoll in Medizin und Biologie zu verwenden.

Ein wichtiger Aspekt hierbei ist die Zuordnung zwischen Gensequenz und Funktion. Heute wird diese Zuordnung einer Funktion zu einer Gensequenz häufig per Analogieschluß durch Übertragung der Funktion einer "ähnlichen" Sequenz (etwa aus einer Sequenzdatenbank) von bekannter Funktionalität vorgenommen. Die grundlegende Idee hierbei ist, daß "ähnliche" Gensequenzen auch einen frühen gemeinsamen Vorfahren (*homolog* zueinander sind) und damit die gleiche Funktion besitzen. Dieser Schluß ist jedoch nur teilweise richtig, da homologe Gene sowohl im Zuge der Artenbildung (*speciation*) als auch durch Genduplikation entstanden sein können.

1. Artenbildung: Spezies A und Spezies B stammen von der gemeinsamen Vorgängerspezies C ab. Das Gen c in Spezies C hat sich im Zuge der Artenbildung zu den homologen Nachfolgenden a in Spezies A und b in Spezies B entwickelt.

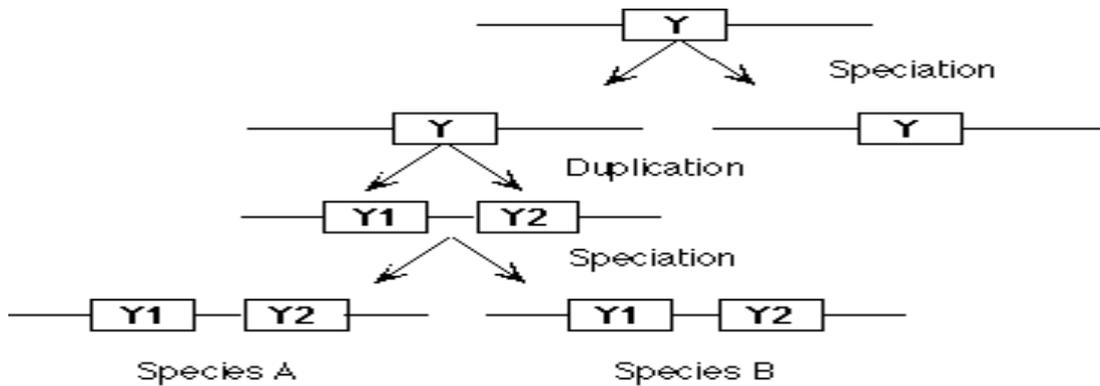


Abbildung 1.1: Paraloge Gene.

2. Duplikation (siehe Abb. 1.1): Eine Genduplikation von Gen *y* in die Kopien *Y1* und *Y2* findet in Spezies *Y* statt. Spezies *A* und *B*, welche sich jeweils aus Spezies *Y* entwickelt haben, beinhalten nun die Nachfolgene beider Kopien: *y1A*, *y2A* in Spezies *A*, *y1B*, *y2B* in Spezies *B*. Die Gene *y1A*, *y2A*, *y1B*, *y2B* sind paarweise zueinander homolog.

Wir bezeichnen homologe Gene, welche im Zuge der Artenbildung entstanden sind, als *orthologe* Gene. Homologe Gene, die durch Duplikation entstanden sind, werden als *paraloge* Gene bezeichnet (Li und Graur, 1991). Damit sind jeweils die Genpaare (*y1A*, *y1B*) und (*y2A*, *y2B*) ortholog, die Genpaare (*y1A*, *y2A*), (*y1B*, *y2B*), (*y1A*, *y2B*) und (*y2A*, *y1B*) paralog zueinander.

Orthologe Gene implizieren, daß die ursprüngliche Funktion des Gens sich unwesentlich in seinen jeweiligen Nachfolgespezies verändert hat, da das Fehlen der alten Funktion sonst zum Aussterben der Spezies geführt hätte. In den meisten Fällen besitzen daher orthologe Gene dieselbe Funktion. Stammt andererseits ein Gen von einer Kopie eines anderen ab, so behält das Nachfolgene der einen Kopie die Funktionalität zur Erhaltung des Organismus bei. Das Nachfolgene der anderen Kopie hat die Freiheit zu mutieren und kann dadurch seine Funktion variieren. Paraloge Gene können, müssen aber nicht dieselbe Funktionalität besitzen. Unter Umständen teilen zwei paraloge Gene einer Multigenfamilie nur einen geringen Anteil an gemeinsamer Funktionalität (Fitch, 1970). Wir lernen also erheblich mehr über Sequenzen und insbesondere deren Funktionalität, wenn wir zusätzlich zu der Homologie die Paralogie oder Orthologie der Sequenzen berücksichtigen.

1.1.2 Rekonstruktion von Genbäumen

Einen Stammbaum für Gene einer Genfamilie zu rekonstruieren, ist ohne Kenntnis der Genduplikationen innerhalb dieser Familie schwierig. Abbildung 1.2 zeigt den vollständigen Genbaum für die Globine der Spezies Mensch, Affe und Pferd. Heute ist bekannt, daß sich die Globine (u. a.) durch ein Duplikationsereignis in ein α - und β -Duplikat aufgeteilt haben. Jeweils aus diesen Duplikaten hat sich dann die Gruppe der α - sowie β -Globine entwickelt.

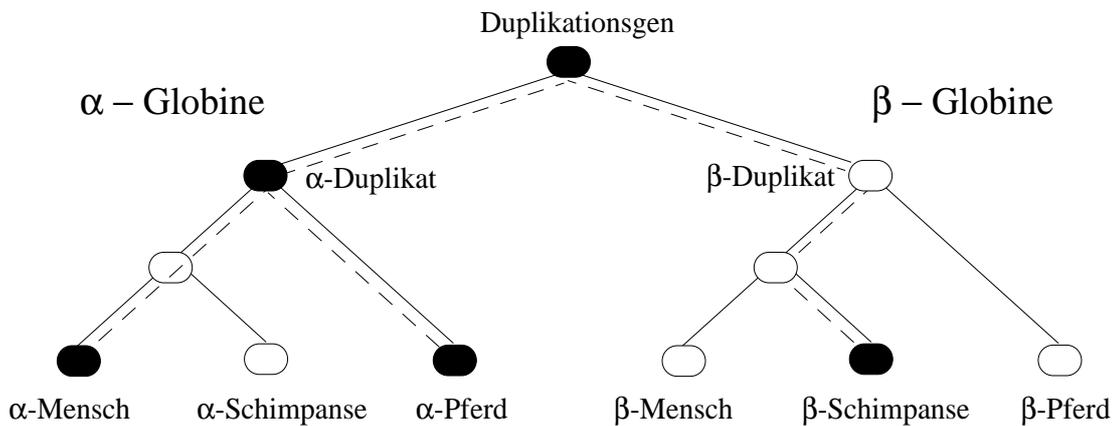


Abbildung 1.2: Vollständiger Genbaum

Rekonstruieren wir nun für die α -Gene von Mensch (α -Mensch) und Pferd (α -Pferd) sowie das β -Gen des Schimpansen (β -Schimpanse) den Genbaum (linker Baum in Abb. 1.3).

Die Inkonsistenz zwischen dem Genbaum und dem heute akzeptierten Speziesbaum (rechter Baum in Abb, 1.3) ergibt sich aufgrund der Genduplikation, durch welche der Genbaum in den sogenannten vollständigen Genbaum einbettbar ist.

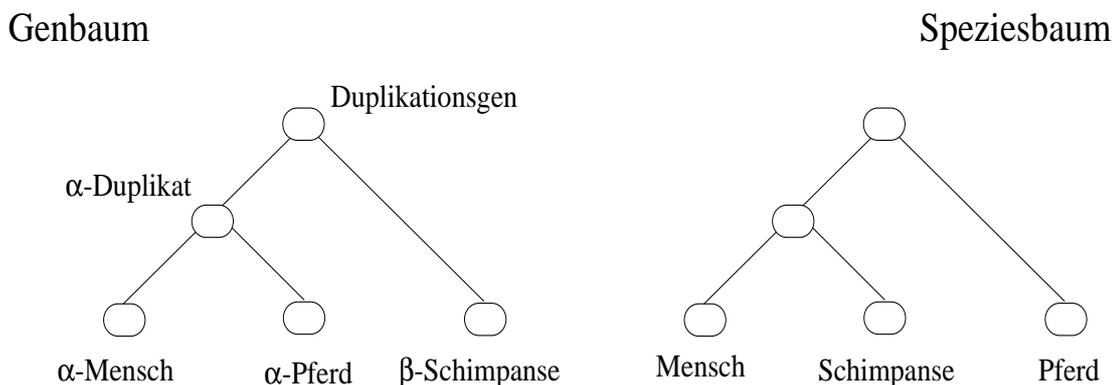


Abbildung 1.3: Inkonsistenz: korrekter Gen- und Speziesbaum

Angenommen, wir hätten das Vorwissen über die Genduplikation nicht, so könnten wir nicht erkennen, daß die Globine α -Mensch und α -Pferd von einer anderen Kopie der Genduplikation abstammen als das Globin β -Schimpanse. Somit wüßten wir nicht, ob die Inkonsistenz zwischen Gen- und Speziesbaum auf eine fehlerhafte Baumrekonstruktion oder auf eine unbekannte Genduplikation zurückzuführen ist. Durch die verschiedenen (2^3) Möglichkeiten, die Globine von Mensch, Pferd und Affe aus der α - oder β -Gruppe der Globine zu wählen, lassen sich alle möglichen (3) Genbäume korrekt rekonstruieren. Die Genduplikation erklärt die Inkonsistenz dieser Genbäume. Das Wissen über die Duplikation einer Genfamilie ermöglicht uns somit, die Inkonsistenz von korrekten Genbäumen zu dieser Familie zu erklären und von solchen Genbäumen zu unterscheiden, deren Inkonsistenz sich möglicherweise durch eine fehlerhafte Rekonstruktion ergeben hat.

Kennen wir hingegen in dem Speziesbaum die Spezies, in welchen eine Genduplikation stattgefunden hat, so können wir aus dem Speziesbaum einen möglichen vollständigen Genbaum ableiten.

Ein Beispiel: Betrachten wir Abbildung 1.4, wobei wir annehmen, daß die Duplikation der Globine in das α - und β -Duplikat in der Wurzel des Speziesbaumes stattgefunden hat. Die Entwicklung der Gene vollzieht sich durch die Entwicklung der Spezies, also entsprechend der Phylogenie des Speziesbaumes. Damit besitzen beide Kopien die Möglichkeit, sich entsprechend der Phylogenie des Speziesbaumes zu entwickeln. Wir erhalten somit den vollständigen Genbaum aus Abbildung 1.2.

Speziesbaum

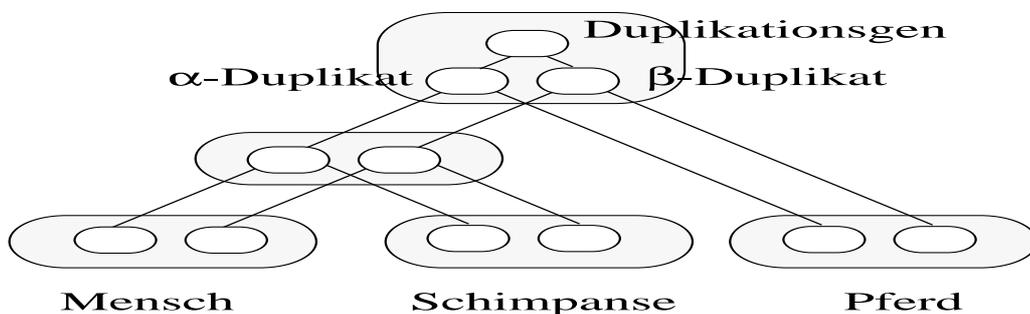


Abbildung 1.4: Entwicklung der Kopien im Speziesbaum

1.2 Die Vorhersage von Genduplikationen

1994 hat der Biologe Page ein biologisches Modell, den *Reconciled Tree*, zu der Vorhersage von Genduplikationen und deren Entwicklung in der Evolution vorgestellt. Ein *Reconciled Tree* ist von Page direkt aus dem Duplikations-Mechanismus der Evolution abgeleitet worden und damit biologisch interpretierbar. Basierend auf einem *Reconciled Tree* hat Page einen Algorithmus entwickelt, welcher unter einer für die Praxis nachteiligen Einschränkung Genduplikationen vorhersagt. Obwohl dieser bereits erfolgreich in der Anwendung eingesetzt wird (in dem Programm COMPONENT), sind dessen Korrektheit und Laufzeit bisher nicht gezeigt worden (eine einfache Analyse zeigt ein quadratisches Laufzeitverhalten). Diese Nachteile liegen im wesentlichen in der indirekten und, für die Konstruktion eines *Reconciled Tree* schwer zugänglichen, semiformalen Beschreibung desselben begründet.

In dieser Arbeit werden wir den *konstruktiven Reconciled Tree* aus einem *Reconciled Tree* von Page ableiten. Der konstruktive *Reconciled Tree* ist zu einem *Reconciled Tree* von Page äquivalent, aber im Gegensatz zu diesem direkt (konstruktiv) und formal definiert. Damit bildet der konstruktive *Reconciled Tree* die Grundlage für alle weiteren Arbeiten. Wir werden Maße zu dem konstruktiven *Reconciled Tree* formulieren, welche Aussagen über die Qualität der vorhergesagten Genduplikationen machen. Basierend auf dem konstruktiven *Reconciled Tree* werden wir einen (korrekten) Algorithmus entwickeln, welcher Genduplikationen vorhersagt und die Qualität dieser Vorhersagen durch Maße zu dem konstruktiven *Reconciled Tree* berechnet. Der Algorithmus hat dabei ein für jede in der Praxis vorkommende Eingabe lineares Laufzeitverhalten. Abschließend werden wir mit dem konstruktiven *Reconciled Tree* Genduplikationen und deren Entwicklung in der Evolution der Farbwahrnehmung bestimmen. Die vorhergesagten Genduplikationen werden in der Evolution der Farbwahrnehmung neue und paraloge Entwicklungen aufzeigen. Diese geben uns dann neue Einsichten in die Entstehung der Farbwahrnehmung.

In den folgenden Abschnitten geben wir einen Überblick zu den einzelnen Kapiteln dieser Arbeit.

1.2.1 Ein *Reconciled Tree* nach Page

Der Biologe Page hat 1994 aufbauend auf den Arbeiten von Goodman et al. (1979) sowie Nelson und Platnick (1981) ein Modell, den *Reconciled Tree*, zu der Vorhersage von Genduplikationen und deren Entwicklung in der Evolution,

vorgestellt (Page, 1994a). Page sagt Genduplikationen und deren evolutionäre Entwicklung für einen gegebenen Genbaum \mathcal{G} und einen gegebenen Speziesbaum \mathcal{S} voraus. Der Genbaum \mathcal{G} beschreibt die Phylogenie einer Menge M von heute existierenden Genen. Die Spezies der Gene, aus welchen die heute existierenden Gene der Menge M stammen, sind bekannt und werden durch dieselbe Menge M beschrieben. Damit identifizieren wir die Menge M der existenten Gene in \mathcal{G} mit den Spezies dieser Gene in dem Speziesbaum \mathcal{S} . Der Speziesbaum \mathcal{S} beschreibt die Phylogenie der Spezies aus der Menge M .

Sind der Genbaum \mathcal{G} und der Speziesbaum \mathcal{S} inkonsistent zueinander, so kann dies durch unerkannte Genduplikationen in \mathcal{G} erklärt werden (siehe Abschnitt 1.1.2). Dazu leitet Page aus dem Speziesbaum \mathcal{S} die an Kardinalität unendliche Menge \mathcal{G} von Genbäumen ab, welche sich durch Genduplikationen ergeben. Im wesentlichen erklärt jeder Genbaum der Menge \mathcal{G} , in welchen sich der Genbaum \mathcal{G} einbetten läßt, die Inkonsistenz zwischen dem Genbaum \mathcal{G} und dem Speziesbaum \mathcal{S} aufgrund von Genduplikationen. Nach dem *Parsimony*-Ansatz wählt Page einen Genbaum \mathcal{R} mit minimaler Anzahl an Knoten aus der Menge \mathcal{G} , in welchen sich der Genbaum \mathcal{G} einbetten läßt. Ein solcher Genbaum wird von Page als Reconciled Tree¹ bezeichnet. Aus dem Reconciled Tree \mathcal{R} lassen sich einerseits die Genduplikationen ablesen, andererseits zeigt er aber auch deren Entwicklung in der Evolution.

Basierend auf dem Modell des Reconciled Tree hat Page einen Algorithmus, den wir als *Page-Algorithmus* bezeichnen, zu der Vorhersage von Genduplikationen und deren evolutionärer Entwicklung in dem Programm COMPONENT 2.0 (1995) integriert (COMPONENT ist ein Programm für die Analyse von evolutionären Bäumen). Der Page-Algorithmus arbeitet zwar auf einem direkt aus der Evolution übertragenen Modell, besitzt jedoch wesentliche Nachteile. Einerseits ist die Korrektheit und Laufzeit des Page-Algorithmus nicht gezeigt worden. Andererseits sagt der Page-Algorithmus Genduplikationen nur für bifurkierende Phylogenien $(\mathcal{G}, \mathcal{S})$ voraus, obwohl ein Reconciled Tree auch für multifurkierende Phylogenien die Vorhersage von Genduplikationen ermöglicht. In der Anwendung ist die Verwendung multifurkierender Phylogenien oft von großer Bedeutung. Häufig entstehen bei der Rekonstruktion des Gen- oder Speziesbaumes Unsicherheiten, welche durch Multifurkationen ausgedrückt werden können. Besitzen zwei Knoten in einem phylogenetischen Baum einen geringen zeitlichen Abstand, so wirkt sich ein kleiner Fehler in der Rekonstruktion dieses Abstands sofort auf die Topologie dieses Baumes aus. Da Genduplikationen auf-

¹Dieser ist der Definition nach nicht eindeutig durch \mathcal{G} und \mathcal{S} bestimmt.

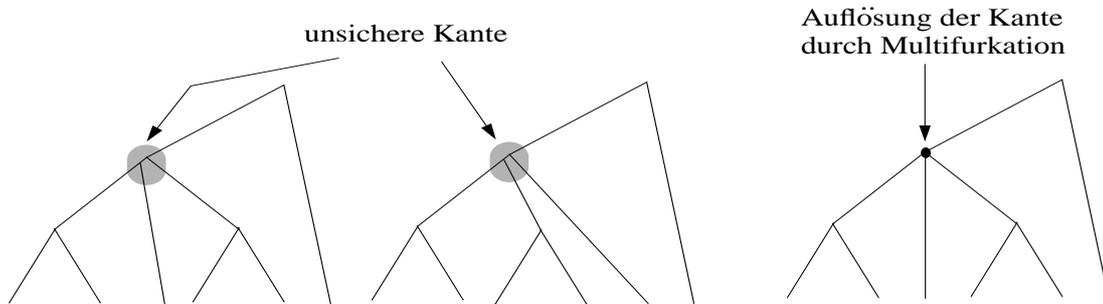


Abbildung 1.5: Die Multifurkation zeigt Unsicherheiten

grund von inkonsistenten Topologien des Gen- und Speziesbaumes vorhergesagt werden, wirkt sich ein solcher Fehler gravierend auf die Qualität der Vorhersage aus. Durch die Verwendung von Multifurkationen kann ein unsicherer kleiner zeitlicher Abstand zwischen zwei Knoten in dem phylogenetischen Baum zu einem Knoten zusammengefaßt werden, wie es z.B. in der Abbildung 1.5 gezeigt ist. Auch die Taxa-Datenbank des NCBI, aus welcher wir in der Anwendung die Speziesbäume ableiten, enthält eine Vielzahl von Multifurkationen (siehe z.B. den Speziesbaum in Abb. 8.4 von Seite 154 des Anwendungskapitels 8). Damit bedeutet in der Anwendung die Einschränkung des Page-Algorithmus auf bifurkierende Phylogenien ein erhöhtes Fehlerpotential in der Vorhersagequalität der Genduplikationen.

Wir werden in Kapitel 3 den Reconciled Tree aus dem Duplikationsmechanismus der Evolution ableiten und dazu die semiformale Beschreibung von Page angeben. Um hieraus Strukturaussagen abzuleiten, geben wir eine formale Definition der Beschreibung von Page an.

1.2.2 Strukturaussagen zu einem Reconciled Tree

Das von Page für die Vorhersage von Genduplikationen und deren Entwicklung verwendete Modell des Reconciled Tree ist direkt von dem Duplikationsmechanismus der Evolution abgeleitet worden. Hieraus ergibt sich auf natürliche Weise eine gute biologische Interpretierbarkeit eines Reconciled Tree. Jedoch ist es schwer, mit diesem Modell die Genduplikationen zu berechnen, da hierzu ein Reconciled Tree in einer an Kardinalität unendlichen Menge von Genbäumen gesucht werden muß. Im einfachsten Ansatz muß jeder dieser Genbäume auf seine Eigenschaft als Reconciled Tree hin untersucht werden. Dabei kann der jeweilige auf den Reconciled Tree zu untersuchende Genbaum eine große Anzahl von Knoten besitzen. Daher verwendet der Page-Algorithmus Strukturaussagen

zu einem Reconciled Tree, mit welchem er die Genduplikationen vorhersagt. Die Korrektheit dieser Strukturaussagen ist jedoch bisher nicht gezeigt worden.

Wir werden in Kapitel 4 wesentliche Strukturaussagen zu einem Reconciled Tree zeigen, mit welchem wir letztendlich dessen eindeutige Topologie beschreiben werden. Somit können wir die Topologie des Reconciled Tree zu einem gegebenen Gen- und Speziesbaum konstruktiv durch Subtopologien des Speziesbaumes beschreiben.

Um die Topologie eines Reconciled Tree zu beschreiben, verwenden wir das Duplikations-Theorem. Dieses beschreibt nur durch den Gen- und Speziesbaum die Genduplikationen in einem zugehörigen Reconciled Tree. Durch das Duplikations-Theorem können wir somit die Genduplikationen vorhersagen, ohne einen Reconciled Tree selbst zu berechnen. Basierend auf dem Duplikations-Theorem werden wir einen Algorithmus konstruieren, welcher die Genduplikationen für einen Reconciled Tree in fast linearer Zeit berechnet. Ohne das Duplikations-Theorem müßten wir einen Reconciled Tree auf Genduplikationen hin durchsuchen. Da dieser, in der Kardinalität des Gen- oder Speziesbaumes, quadratisch viele Elemente besitzen kann, würde ein Algorithmus dazu mindestens ein quadratisches Laufzeitverhalten besitzen.

Als Nebeneffekt sind die von dem Page-Algorithmus verwendeten Strukturaussagen in den von uns gezeigten Strukturaussagen enthalten, womit die Korrektheit des Page-Algorithmus im wesentlichen gezeigt ist.

1.2.3 Der konstruktive Reconciled Tree

Mit den zuvor aus dem Reconciled Tree abgeleiteten Strukturaussagen werden wir in Kapitel 5 den konstruktiven Reconciled Tree definieren. Der konstruktive Reconciled Tree zu einem Genbaum \mathcal{G} und einem Speziesbaum \mathcal{S} besitzt dieselbe Topologie wie ein Reconciled Tree zu \mathcal{G} und \mathcal{S} . Aufgrund derselben Topologie ist der konstruktive Reconciled Tree ein zu dem Reconciled Tree äquivalentes Modell, um Genduplikationen und deren Entwicklung in der Evolution vorherzusagen.

Die Definition des konstruktiven Reconciled Tree gibt uns, im Gegensatz zu einem Reconciled Tree, eine konkrete und eindeutige Vorstellung von dessen Topologie. Denn diese setzt sich aus – in der Definition durch die Topologie von \mathcal{G} direkt angegebenen – Subtopologien des Speziesbaumes \mathcal{S} zusammen. Damit ist der konstruktive Reconciled Tree auch direkt, aus seiner Definition heraus, konstruierbar. Die Benennung der Elemente des konstruktiven Reconciled Tree ist so gewählt worden, daß diese eine direkte biologische Interpretation besitzen und sich auch formal einfach handhaben lassen.

1.2.4 Maße zum konstruktiven Reconciled Tree

Um eine Aussage über die Qualität von vorhergesagten Genduplikationen zu erhalten, führen wir in Kapitel 6 Maße zu dem konstruktiven Reconciled Tree ein. Dazu beginnen wir mit der Definition eines *groben* Maßes, welches nach einem biologischen Konzept die Qualität von Genduplikationen des konstruktiven Reconciled Tree bewertet. Aus dem *groben* Maß definieren wir dann durch schrittweise Verfeinerung des biologischen Konzeptes weitere, *sensiblere*, Maße. Die so definierten Maße basieren damit auf Verfeinerungen desselben biologischen Konzeptes.

Wir zeigen dann, daß im wesentlichen alle aus der Literatur bekannten Maße äquivalent zu den von uns definierten Maßen sind, wenn wir unsere Maße auf bifurkierende Phylogenien beschränken. Somit können die aus der Literatur bekannten Maße nach demselben biologischen Konzept abgeleitet werden und Äquivalenzen zwischen in der Literatur unterschiedlich definierten Maßen aufgedeckt werden.

Bemerkenswert ist das Maß M_{MMS} , welches sich auch nach dem von uns verwendeten Konzept ableiten läßt. Denn dieses Maß ist nach einem biologischen Konzept definiert, welches sich wesentlich von dem unsrigen unterscheidet. Erst die *Mirkin-Muchnik-Smith-Vermutung* (Mirkin et al., 1995), bekannt als *MMS-Vermutung*, gab uns den Hinweis darauf, auch dieses Maß nach dem von uns verwendeten Konzept ableiten zu können. Die MMS-Vermutung besagt, daß das Maß M_{MMS} zu einem nach unserem Konzept ableitbaren Maß $M_{Muchnik}$ äquivalent ist. Der Beweis dieser Vermutung (Eulenstein und Vingron, 1995) zeigte dann, daß zwei gänzlich unterschiedliche biologische Konzepte zu demselben Maß führen.² Wir werden in dem Kapitel 6 die grundlegende Idee des Beweises von Eulenstein und Vingron aufzeigen.

Damit lassen sich alle in dieser Arbeit aus der Literatur angegebenen Maße durch ein von uns definiertes Maß berechnen. Im wesentlichen läßt sich jedes unserer Maße in linearer Zeit berechnen und damit jedes uns aus der Literatur bekannte Maß.

²Der Beweis der MMS- Vermutung ist erstmals von Eulenstein und Vingron (1995) gezeigt worden. Von diesem Beweis unabhängig ist die Vermutung dann von Zhang (1997) bestätigt worden. Varianten des ursprünglichen Beweises von Eulenstein und Vingron sind in Eulenstein, Mirkin und Vingron (1997) sowie Eulenstein und Vingron (1998) zu finden.

1.2.5 Algorithmus

In dieser Arbeit werden wir in Kapitel 7 erstmals die Korrektheit eines Algorithmus zeigen, welcher Genduplikationen auf multifurkierenden Phylogenien vorhersagt und die Qualität dieser Vorhersage durch ein biologisch sensibles Maß bewertet. Dieser Algorithmus besitzt dazu ein für alle in der Praxis vorkommenden Anwendungen lineares Laufzeitverhalten.

Mit den vorhergesagten Genduplikationen kann dann der konstruktive Reconciled Tree direkt nach seiner Definition und damit mit einer linearen Laufzeit in der Anzahl seiner Elemente konstruiert werden. Der konstruktive Reconciled Tree zu einem Genbaum \mathcal{G} und einem Speziesbaum \mathcal{S} besitzt im schlechtesten Fall quadratisch viele Elemente in der Größe des Speziesbaumes \mathcal{S} . Somit läßt sich der konstruktive Reconciled Tree in quadratischer Zeit in der Größe des Speziesbaumes \mathcal{S} berechnen.

Wir werden außerdem eine einfache Reduktion des Problemes der Vorhersage von Genduplikationen zeigen, welche durch das Duplikations-Theorem ermöglicht wird. Durch die Reduktion erhalten wir asymptotisch dieselbe Laufzeit, wie beim zuerst genannten maßgeschneiderten Algorithmus. In der Anwendung ist der maßgeschneiderte Algorithmus jedoch schneller als ein Algorithmus, welcher auf der Reduktion basiert.

1.2.6 Anwendung

Die Biologen Kawamura und Yokoyama (1995) haben aufgrund einer neu sequenzierten Sequenz eine Genduplikation in der Evolution der Farbwahrnehmung postuliert. Wie Kawamura und Yokoyama werden auch wir mit dem konstruktiven Reconciled Tree in Kapitel 8 Genduplikationen in der Evolution der Farbwahrnehmung vorhersagen.

Eine Analyse mit dem Reconciled Tree von Page ist zuvor durch Yuan, Eulenstein, Vingron und Bork (1998) erfolgt. Zu diesem Zeitpunkt war die Vorhersage von Genduplikationen jedoch noch auf bifurkierende Phylogenien beschränkt und nicht gezeigt, daß der dort analysierte Reconciled Tree der Definition von Page entspricht. Aufgrund der Verwendung des eingeschränkten Modells des Reconciled Tree konnten vorhandene Ungenauigkeiten im Speziesbaum nicht ausgedrückt und dadurch in den vorhergesagten Genduplikationen und deren Entwicklung nicht mehr aufgezeigt werden.

In der hier gezeigten Analyse werden wir die Genduplikationen erstmals für multifurkierende Phylogenien vorhersagen und an dem zugehörigen konstruktiven Reconciled Tree die evolutionäre Entwicklung der Genduplikationen analysieren.

Kapitel 2

Darstellung von phylogenetischen Bäumen

Das grundlegende Objekt unserer Betrachtung ist der phylogenetische Baum in seinen unterschiedlichen Bedeutungen wie Gen- oder Speziesbaum. Üblicherweise werden gewurzelte phylogenetische Bäume in der systematischen Biologie als Mengensysteme dargestellt. Auch die in dieser Arbeit verwendeten Aussagen aus der Literatur stellen den phylogenetischen Baum als Mengensystem dar. Diese Darstellung werden wir jedoch nicht alleine aufgrund ihrer allgemeinen Akzeptanz verwenden. Daher werden wir in Abschnitt 2.1 die Mengensystemdarstellung direkt aus den Anforderungen dieser Arbeit an den phylogenetischen Baum ableiten und diese dann in Abschnitt 2.1 formal definieren.

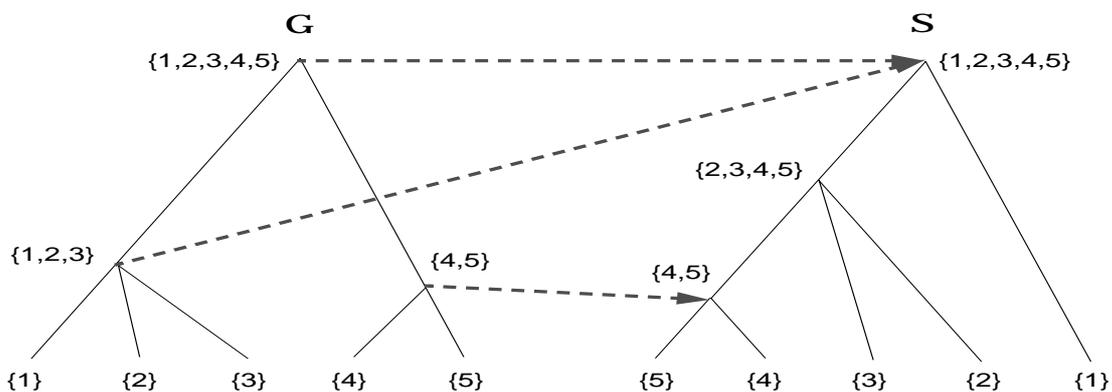


Abbildung 2.1: Die Hasse-Diagramme der Mengensysteme von \mathcal{G} und \mathcal{S}

2.1 Die Darstellung als Mengensystem

Die wahrscheinlich intuitiv eingängigste Darstellung des phylogenetischen Baumes ist die des Baumgraphen. Als Beispiel sind in Abbildung 2.1 zwei phylogenetische Bäume, ein Genbaum \mathcal{G} und ein Speziesbaum \mathcal{S} , als Baumgraphen dargestellt. Innere Knoten des Genbaumes \mathcal{G} sind Gene, welche früher einmal existierten. Aus diesen Genen haben sich, entsprechend der Topologie des Genbaumes \mathcal{G} , die heute existierenden Gene, die Blätter von \mathcal{G} , entwickelt. Analoges gilt für den Speziesbaum \mathcal{S} .

Da wir die früher existenten Gene und Spezies meist nicht kennen, werden diesen künstliche Bezeichnungen zugeordnet. Üblicherweise werden diese Gene oder Spezies durch Mengen, wie in Abbildung 2.1 gezeigt, bezeichnet. Diese Bezeichnungen erhält man, indem auch den Blättern des Gen- und Speziesbaumes Mengen zugeordnet werden. Um einen Knoten X eines Gen- oder Speziesbaumes durch eine Menge zu bezeichnen, betrachtet man die Blätter des Teilbaumes, welcher durch X gewurzelt ist. Durch die Vereinigung der Mengen dieser Blätter wird dann der Knoten X gekennzeichnet.

Im wesentlichen werden wir einen Genbaum \mathcal{G} und einen Speziesbaum \mathcal{S} gemeinsam betrachten. Dies liegt daran, daß uns die Spezies des Speziesbaumes \mathcal{S} interessieren, welche ein Gen des Genbaumes \mathcal{G} beinhalten können. Hierunter verstehen wir, daß ein Organismus einer Spezies existiert hat, welcher dieses Gen aufgrund der Phylogenien \mathcal{G} und \mathcal{S} beinhaltet haben kann.

Ein Gen $X \in \mathcal{G}$ kann in einer Spezies $X' \in \mathcal{S}$ genau dann enthalten sein, wenn die Spezies des Teilbaumes in \mathcal{S} gewurzelt an X' alle Gene, des Teilbaumes in \mathcal{G} gewurzelt an X beinhalten.

Ein Beispiel: Betrachten wir in Abbildung 2.1 das Gen $\{4, 5\}$. Setzen wir voraus, daß wir die Spezies der heute existierenden Gene kennen und die Gene der Blätter von \mathcal{G} in den identisch bezeichneten Spezies beinhaltet sind. Dann kann das Gen $\{4, 5\}$ nur in den Spezies $\{4, 5\}$, $\{2, 3, 4, 5\}$ und $\{1, 2, 3, 4, 5\}$ beinhaltet sein.

Die von heute an betrachtete jüngste Spezies $X' \in \mathcal{S}$, die ein Gen X beinhaltet haben kann, bezeichnen wir als *die Host-Spezies von X'* .

In Abbildung 2.1 ist z.B. die Spezies $\{4, 5\}$ die Host-Spezies des Gens $\{4, 5\}$. Da einem Molekularbiologen die Host-Spezies, aus denen er seine Gene extrahiert,

bekannt sein sollten, setzen wir diese als bekannt voraus. Die Bezeichner dieser Gene identifizieren wir allgemein mit den Bezeichnern ihrer Host-Spezies und leiten die folgende Voraussetzung ab.

Die Bezeichner der Blätter in \mathcal{G} sind identisch mit denen der Blätter in \mathcal{S} .¹

Mit dieser Voraussetzung können wir nun jedem Gen in \mathcal{G} seine Host-Spezies zuordnen. Ein Beispiel: Wir wollen Gen $\{1, 2, 3\} \in \mathcal{G}$ in Abbildung 2.1 seine Host-Spezies zuordnen. Die Nachfolge-Gene von $\{1, 2, 3\}$ sind die Gene $\{1\}$, $\{2\}$ und $\{3\}$, welche nach Voraussetzung in den gleichnamigen Spezies beinhaltet sind. Damit ist die Spezies $\{1, 2, 3, 4, 5\}$ die einzige, in der das Gen $\{1, 2, 3\}$ beinhaltet sein kann. Somit ist Spezies $\{1, 2, 3, 4, 5\}$ auch die Host-Spezies des Gens $\{1, 2, 3\}$.

Ordnen wir also einem Gen $X \in \mathcal{G}$ seine Host-Spezies zu, so betrachten wir die Blätter des Teilbaumes von \mathcal{G} gewurzelt an X . Diese Blätter identifizieren wir mit den Namen ihrer entsprechenden Host-Spezies und suchen für diese Spezies den kleinsten Teilbaum in \mathcal{S} , welcher nur diese als Blätter beinhaltet. Die Wurzel dieses Teilbaumes ist dann die Host-Spezies von X .

Den Vorgang, die Blätter eines Teilbaumes, welcher durch ein Gen oder eine Spezies gewurzelt ist, zu suchen, verkürzen wir. Dazu bezeichnen wir ein Gen oder eine Spezies durch die Menge der Blätter des Teilbaumes, welcher durch dieses Gen oder diese Spezies gewurzelt ist (wie wir es schon in Abbildung 2.1 getan haben). Also ist X' die Host-Spezies des Gens X , wenn $X \subseteq X'$ gilt und kein $Y' \in \mathcal{S}$ mit $X \subseteq Y' \subset X'$ existiert. Dies entspricht der kleinsten oberen Schranke von X in den Mengenbezeichnern von \mathcal{S} .

Die Mengen-Bezeichner der Knoten eines phylogenetischen Baumes bilden ein spezielles Mengensystem, dessen Hasse-Diagramm isomorph zu dem phylogenetischen Baumgraphen ist. Ein solches Mengensystem beschreibt also einerseits durch sein Hasse-Diagramm den phylogenetischen Baum und stellt andererseits die auf Mengen üblichen Relationen zur Verfügung, mit welchen wir die Host-Spezies beschreiben. Estabrook und McMorris (1980) haben ein Mengensystem mit den hier beschriebenen Eigenschaften als n -Clustersystem definiert.

¹Damit kann \mathcal{G} keine Blätter derselben Host-Spezies besitzen, welches ein Nachteil in der Anwendung wäre. Vingron (persönlich kommuniziert) hat jedoch für Blätter des Genbaumes derselben Host-Spezies jeweils neue Pseudo-Spezies eingeführt und damit den Nachteil in der Anwendung beseitigt.

2.2 Das n-Clustersystem

In diesem Abschnitt führen wir die mengentheoretische Darstellung des phylogenetischen Baumes von McMorris und Estabrook (1980), das n-Clustersystem, ein. Wir zeigen dann, daß das n-Clustersystem eine Darstellung für den phylogenetischen Baum ist.

Das n-Clustersystem ist eine Menge von Bezeichnern eines phylogenetischen Baumes mit gewissen Eigenschaften. Damit ein phylogenetischer Baum durch eine Menge dargestellt werden kann, muß diese Menge außer den Bezeichnern auch die Topologie des phylogenetischen Baumes wiedergeben. In Abschnitt 2.2.1 zeigen wir, daß diese Topologie allgemein durch eine bestimmte Ordnung, die phylogenetische Baumordnung, auf einer Menge dargestellt werden kann. In Abschnitt 2.2.2 geben wir dann die Definition des n-Clustersystems von Estabrook und McMorris (1980) an und zeigen, daß dieses eine Baumordnung besitzt. Damit ist das n-Clustersystem eine Darstellung für den phylogenetischen Baum.

2.2.1 Die Baumordnung

Damit ein phylogenetischer Baum durch eine Menge darstellbar ist, muß diese eine zu dem Baumgraphen äquivalente Topologie besitzen. Die Knoten des Baumgraphen werden durch eine Menge dargestellt und die Kanten zwischen den Knoten durch eine Relation, welche auf dieser Menge eine Baumordnung ist.

Definition 2.1 (Baumordnung)

Sei M eine Menge und \sqsubseteq eine Relation auf dieser Menge. Wir bezeichnen diese Relation als Baumordnung auf M , wenn die folgenden Aussagen gelten (die binären Relationen \sqcup, \sqcap und \diamond sind im Appendix Def. 9.5 sowie Def. 9.4 definiert).

1. \sqsubseteq ist auf M eine Ordnungsrelation.
2. M ist gegen \sqcup abgeschlossen.
3. $\forall X, Y \in M : \neg(X \diamond Y) \Rightarrow X \sqcap Y \notin M$.

Für eine Menge M mit einer Baumordnung führen wir nun einige Bezeichnungen ein, welche sich auf natürliche Weise von den Baumgraphen ableiten.

Definition 2.2 (Wurzel, Blätter, Kinder, Eltern, Teilbaum)

Sei M eine Menge mit einer Baumordnung \sqsubseteq und $x, y \in M$.

- Wurzel: $\text{Ro}(M) := \sqcup M$,
bezeichnen wir als Wurzel von M (Ro für Root).
 Blatt: $\text{Le}(M) := \{l \in M \mid \nexists s \in M : s \sqsubset l\}$,
bezeichnen wir als die Blätter von M (Le für Leaves).
 Kinder: $\text{Ch}(x) := \{z \in M \mid z \sqsubset x \wedge \nexists y \in M : y \neq z \Rightarrow z \sqsubset y \sqsubset x\}$,
bezeichnen wir als die Kinder von x (Ch für Children).
 Eltern: $\text{Pa}(z) = x : \Leftrightarrow z \in \text{Ch}(x)$,
bezeichnen wir als Eltern von x (Pa für Parent).
 Teilbaum: $\text{T}_M(x) := \{y \in M \mid y \sqsubseteq x\}$,
bezeichnen wir als Teilbaum von M gewurzelt an x.

Das Äquivalent des Weges in einem Baumgraphen von x nach y der Baumordnung ist die Kette $K(X, Y)$.

Mit der Baumordnung läßt sich allgemein ein Baumgraph beschreiben. Damit ein Baumgraph einen phylogenetischen Baum darstellt, muß er noch die beiden folgenden Eigenschaften erfüllen.

- Die Menge der Knoten des Baumgraphen ist endlich.
- Kein Knoten eines Baumgraphen hat genau ein Kind.

Definition 2.3 (Phylogenetische Baumordnung)

Sei M eine endliche Menge mit einer Relation \sqsubseteq . Wir bezeichnen diese Relation als phylogenetische Baumordnung auf M , wenn \sqsubseteq auf M eine Baumordnung ist und $\forall x \in M : |\text{Ch}(x)| \neq 1$ gilt.

2.2.2 Das n-Clustersystem

Wir werden die Definition des n-Clustersystemes angeben, welches wir in Abschnitt 2.1 vom phylogenetischen Baumgraphen abgeleitet haben und das von Estabrook und McMorris (1980) definiert wurde. Damit das n-Clustersystem eine zu dem phylogenetischen Baum äquivalente Darstellung ist, zeigen wir dann, daß dieses mit der Relation \subseteq eine phylogenetische Baumordnung besitzt.

Definition 2.4 (n-Clustersystem)

Sei M eine endliche Menge und $\mathcal{M} \subset \wp(M)$. Wir bezeichnen \mathcal{M} als n-Clustersystem über M , wenn es folgenden Bedingungen genügt:

- | | |
|---|--|
| (i) $M \in \mathcal{M}$ | (ii) $\emptyset \notin \mathcal{M}$ |
| (iii) $\forall i \in M : \{i\} \in \mathcal{M}$ | (iv) $\forall X, Y \in \mathcal{M} : X \cap Y \in \{\emptyset, X, Y\}$ |

Die Menge aller n-Clustersysteme über M bezeichnen wir mit $\mathfrak{N}(M)$.

Lemma 2.1

Sei M eine Menge und $\mathcal{M} \in \mathfrak{N}(M)$, dann ist \subseteq ein Baumordnung auf \mathcal{M} .
Wir zeigen dazu:

Beweis.

1. \subseteq ist Ordnung auf \mathcal{M} : Dies gilt offensichtlich.
2. \mathcal{M} ist gegen \sqcup abgeschlossen: Seien $X, Y \in \mathcal{M}$. Da $\mathcal{M} \in \mathfrak{N}(M)$ gilt, folgt $X \subseteq M$ und $Y \subseteq M$. Nach Def. 2.4(1) ist $M \in \mathcal{M}$. Also ist M eine obere Schranke für X und Y . \mathcal{M} ist endlich, da M endlich ist. Somit existiert $X \sqcup Y$ in \mathcal{M} .
3. $\forall X, Y \in \mathcal{M} : \neg(X \diamond Y) \Rightarrow X \sqcap Y \notin \mathcal{M}$.
Seien $X, Y \in \mathcal{M}$ mit $\neg(X \diamond Y)$. Mit Bedingung (iv) der Def. 2.4 folgt $X \cap Y = \emptyset$ hieraus. Damit ist $X \sqcap Y = \emptyset$. Nach Bedingung (ii) der Def. 2.4 gilt $\emptyset \notin \mathcal{M}$. Somit gilt $X \sqcap Y \notin \mathcal{M}$.
4. $\forall X \in \mathcal{M} : |\text{Ch}(X)| \neq 1$. Annahme: $\exists X \in \mathcal{M} : |\text{Ch}(X)| = 1$.
Sei $\text{Ch}(X) = \{Z\}$. Damit gilt $Z \subset X$. Wegen $Z \subset X$ existiert ein $i \in X \setminus Z$. Da $\mathcal{M} \in \wp(M)$, ist $X \subseteq M$, womit $i \in M$ ist. Nach Bedingung (iii) der Def. 2.4 ist $I := \{i\} \in \mathcal{M}$. Also existiert ein $I \in \mathcal{M}$ mit $I \not\subseteq Z$ und $I \subset X$, womit ein $Y \in \text{Ch}(X)$ mit $Z \neq Y$ existiert. Dies ist ein Widerspruch zu der Annahme.

□

Damit kann der phylogenetische Baum durch ein n -Clustersystem dargestellt werden. Wenn wir im weiteren von einem phylogenetischen Baum sprechen, so ist hierunter dessen Darstellung als n -Clustersystem zu verstehen. Außerdem betrachten wir von nun an nur noch endliche Mengen.

2.3 Die Erweiterung der kleinsten oberen Schranke

Sei M eine Menge und $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$, wobei \mathcal{G} ein Genbaum und \mathcal{S} ein Speziesbaum ist. Wir werden nun, wie im vorigen Abschnitt beschrieben, einem Gen $X \in \mathcal{G}$ seine Host-Spezies $X' \in \mathcal{S}$ durch die kleinste obere Schranke der Elemente aus X in \mathcal{S} zuordnen. Hierzu erweitern wir die Definition 9.5 der kleinsten oberen Schranke.²

²Die kleinste obere Schranke $\sqcup K$ in einer Menge M ist nur für Elemente $K \subseteq M$ definiert. Für die Bestimmung der Host-Spezies benötigen wir jedoch die kleinste obere Schranke $\sqcup K$ in der Menge \mathcal{S} mit $K \in \mathcal{G}$.

Definition 2.5 ($\bigsqcup_{\mathcal{T}}(X)$ kleinste obere Schranke von X in \mathcal{T})

Sei M eine Menge, $\mathcal{M} \in \mathfrak{N}(M)$ und $X \subseteq M$. Dann bezeichnen wir

$$\bigsqcup_{\mathcal{M}}(X) := \bigsqcup \{ \{i\} \in \mathcal{M} \mid i \in X \}$$

als die kleinste obere Schranke von X in \mathcal{M} .

Seien $A, B \in \mathcal{G}$. Es gilt offensichtlich

$$A \subseteq \bigsqcup_{\mathcal{S}}(A), \quad A \subseteq B \Rightarrow \bigsqcup_{\mathcal{S}}(A) \subseteq \bigsqcup_{\mathcal{S}}(B)$$

Im allgemeinen gilt nicht

- (i) $\bigsqcup_{\mathcal{S}}(A \sqcup_{\mathcal{G}} B) = \bigsqcup_{\mathcal{S}}(A) \sqcup_{\mathcal{S}} \bigsqcup_{\mathcal{S}}(B)$
- (ii) $A \subset B \Rightarrow \bigsqcup_{\mathcal{S}}(A) \subset \bigsqcup_{\mathcal{S}}(B)$

Wir zeigen die Gegenbeispiele für \mathcal{G} und \mathcal{S} der jeweils angegebenen Abbildungen. Die Pfeile der Abbildung geben für jeden inneren Knoten $X \in \mathcal{G}$ die Menge $\bigsqcup_{\mathcal{S}}(X) \in \mathcal{S}$ an.

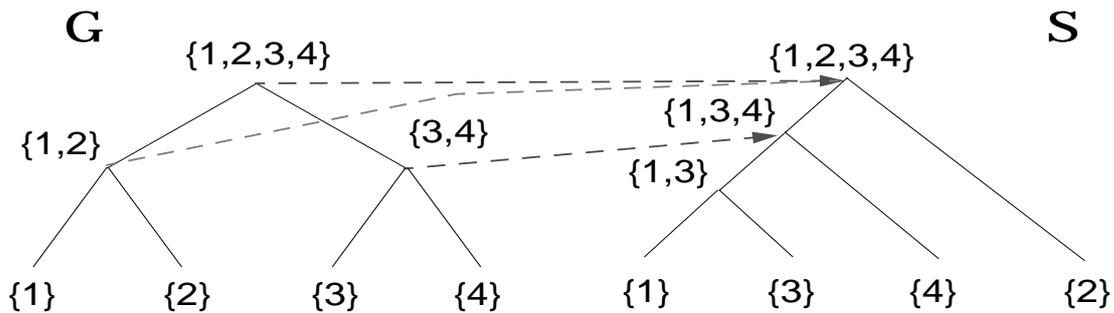


Abbildung 2.2: Ein Genbaum \mathcal{G} und Speziesbaum \mathcal{S} .

Zu (i): Seien $\{1\}, \{2\} \in \mathcal{G}$, dann gilt in Abb. 2.1:

$$\{1\} \sqcup_{\mathcal{G}} \{2\} = \{1, 2, 3\} \neq \{1, 2, 3, 4, 5\} = \bigsqcup_{\mathcal{S}}(\{1\}) \sqcup_{\mathcal{S}} \bigsqcup_{\mathcal{S}}(\{2\})$$

Zu (ii): Seien $\{1, 2\}, \{1, 2, 3, 4\} \in \mathcal{G}$, dann gilt in Abb. 2.2:

$$\{1, 2\} \subset \{1, 2, 3, 4\} \wedge \bigsqcup_{\mathcal{S}}(\{1, 2\}) = \{1, 2, 3, 4\} = \bigsqcup_{\mathcal{S}}(\{1, 2, 3, 4\})$$

Kapitel 3

Der Reconciled Tree nach R. Page

Der Reconciled Tree ist ein spezieller Genbaum, welcher die sensitive Vorhersage von Genduplikationen und deren Entwicklung in der Evolution durch den Rechner ermöglicht. Das Modell des Reconciled Tree ist von dem Biologen Page (1994a), aufbauend auf den Arbeiten von Goodman et al. (1979) sowie Nelson und Platnick (1981), entwickelt und in den Arbeiten von Page (1994a), Slowinski et al. (1997), Page et al. (1997) und Yuan et al. (1998) als das zentrale Modell zu der Analyse von Genduplikationen und deren Entwicklung eingesetzt worden.

In diesem Kapitel werden wir den Reconciled Tree von Page einführen und hierzu eine formale Definition angeben. Aus der formalen Definition werden wir dann in dem folgenden Kapitel 4 strukturelle Aussagen ableiten, mit denen wir den Reconciled Tree schnell berechnen können.

Um den Reconciled Tree einzuführen, zeigen wir in Abschnitt 3.1, daß Page den Reconciled Tree auf natürliche Weise und direkt aus der Evolution abgeleitet hat. Basierend hierauf hat Page dann eine semiformale Definition des Reconciled Tree angegeben (Page,1994a), die wir in Abschnitt 3.2 zeigen werden, um daraus abschließend in Abschnitt 3.2 eine formale Definition des Reconciled Tree von Page anzugeben.

3.1 Die Ableitung aus der Evolution

Angenommen, der Genbaum \mathcal{G} in Abbildung 3.1 ist für einen Gentyp korrekt rekonstruiert worden. Es stimmt mit unserer Vorstellung von der Speziesentwicklung überein, daß das Gen 4 des Affen und das Gen 5 des Menschen einen

früheren gemeinsamen Vorfahren besitzen als diese mit jedem anderen Gen (1, 2, und 3). Aber, daß das Gen 2 des Pferdes und das Gen 3 des Vogels einen früheren gemeinsamen Vorfahren besitzen als Gen 2 des Pferdes und Gen 4 des Affen, widerspricht unserer Vorstellung.

Die grobe Vorstellung davon, wie der Genbaum aussehen sollte, leiten wir direkt von der Entwicklung der Spezies ab. In gewisser Weise ist dies korrekt, da sich die Entwicklung der Gene durch die Entwicklung der Spezies vollzieht. Wir sollten demnach eine Entwicklung der Gene in \mathcal{G} erwarten, welche konsistent zu der Entwicklung der Spezies ist. Aus \mathcal{S} leiten wir also einen Genbaum ab, welcher konsistent zu \mathcal{S} ist. \mathcal{G} ist jedoch nicht auf diese Weise aus \mathcal{S} ableitbar.

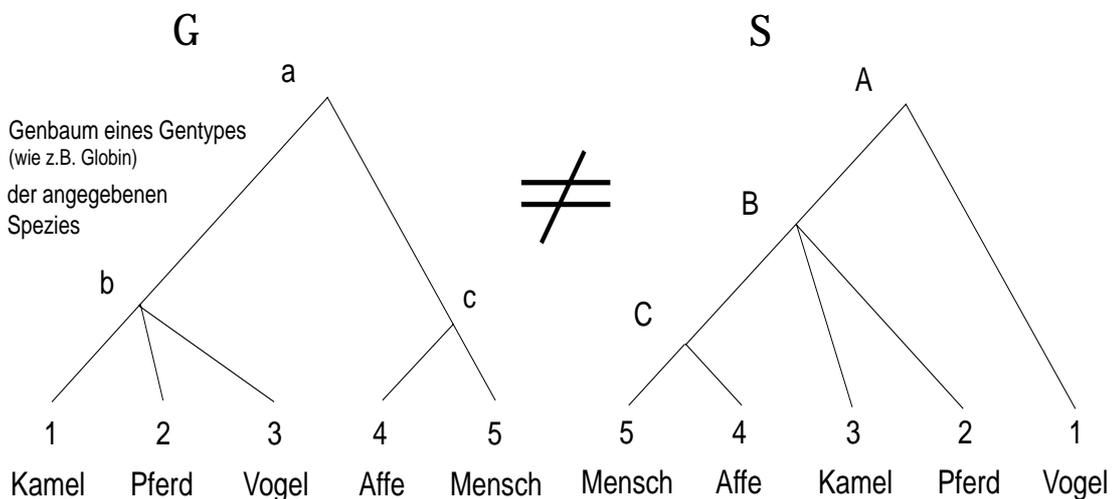


Abbildung 3.1: Inkonsistenz eines korrekten Genbaumes \mathcal{G} und Speziesbaumes \mathcal{S}

In der Evolution existieren Mechanismen, durch welche eine zu \mathcal{S} inkonsistente Entwicklung der Gene entsteht. Wichtige Mechanismen sind dabei der horizontale Gen-Transfer und die Genduplikation.

Der horizontale Gen-Transfer: Bisher haben wir Gene betrachtet, welche sich von einer Spezies zu deren Nachfolgespezies weiterentwickeln. Jedoch kann ein Gen, beispielsweise durch Parasiten, auch von einer Spezies eines Teilbaumes von \mathcal{S} zu einer Spezies eines vollkommen anderen Teilbaumes in \mathcal{S} transferiert werden. Ein Gen ist dabei nur zwischen existenten Spezies – und damit horizontal in \mathcal{S} – transferierbar. Somit kann ein horizontal zwischen Spezies transferiertes Gen die Inkonsistenz zwischen \mathcal{G} und \mathcal{S} verursachen. Page (1994b) hat die Inkonsistenz von Gen und Speziesbaum aufgrund des horizontalen Gen-Transfers untersucht.

In dieser Arbeit beschränken wir uns auf die Genduplikation und setzen damit voraus, daß sich die Inkonsistenz zwischen \mathcal{G} und \mathcal{S} nur durch die Genduplikation erklärt.

Die Genduplikation: Bei der Genduplikation wird ein Gen in einer Spezies dupliziert. Ein solches Gen bezeichnen wir als *Duplikationsgen*. Findet keine weitere Genduplikation statt, entwickelt sich jede Kopie des Duplikationsgens konsistent zu \mathcal{S} weiter.

Wir zeigen, wie der zu \mathcal{S} inkonsistente Genbaum \mathcal{G} durch Gen-Duplikationen aus \mathcal{S} ableitbar ist. Betrachten wir dazu in Abbildung 3.2 das Gen A_0 in Spezies A. A_0 ist ein Duplikationsgen, dessen Kopien die Gene A_1 und A_2 in Spezies A sind. A_1 ist wieder ein Duplikationsgen mit den Kopien A_{11} , A_{12} und A_{13} in der Spezies A. Somit können sich die Gene A_{11} , A_{12} , A_{13} und A_2 konsistent zu \mathcal{S} weiterentwickeln, wie es in Abbildung 3.2 dargestellt ist.

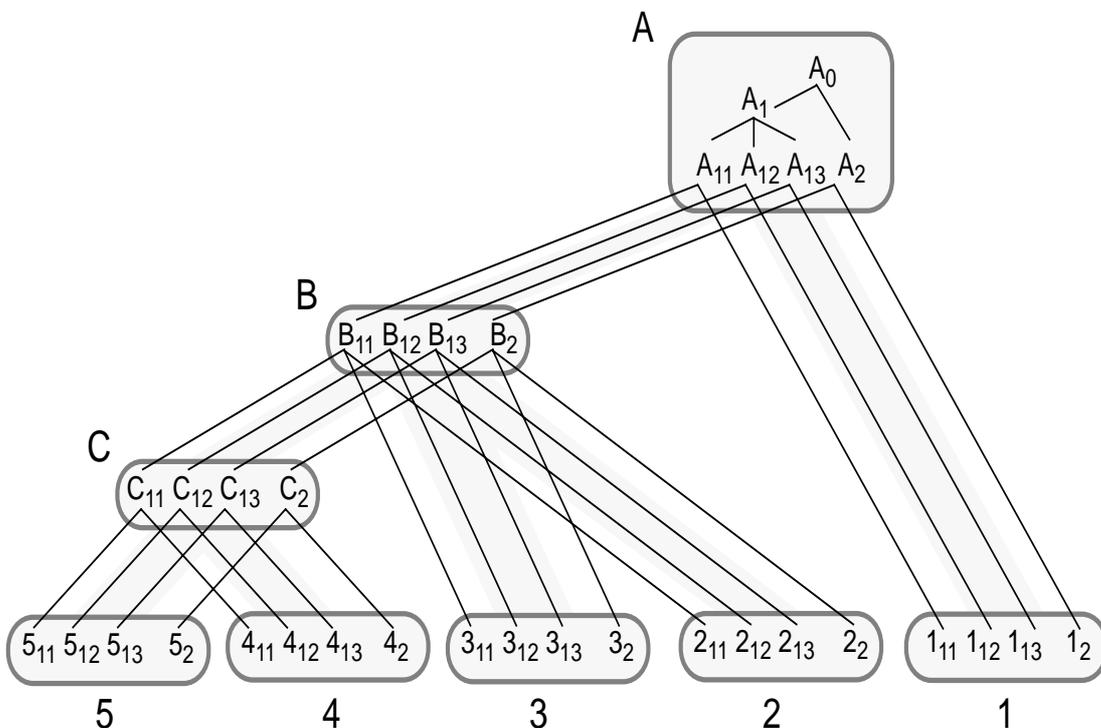


Abbildung 3.2: Die Ableitung \mathcal{R} eines zu \mathcal{S} konsistenten Genbaumes

Damit der abgeleitete Genbaum \mathcal{R} den Genbaum \mathcal{G} erklärt, muß er eine Einbettung von \mathcal{G} beinhalten, die spezieskonform ist. *Spezieskonform* bedeutet, daß jedes Gen $X \in \mathcal{G}$ in der Spezies des Gens $X' \in \mathcal{R}$, in welches X eingebettet ist, enthalten sein kann. Also ist die Host-Spezies von X gleich oder entwickelt sich aus der Spezies, in welcher das Gen X' enthalten sein kann.

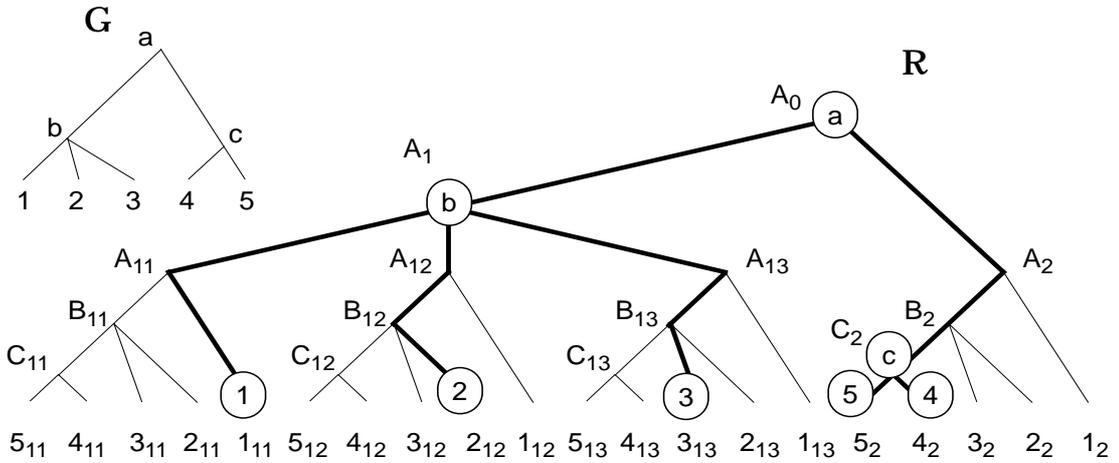


Abbildung 3.3: Eine spezieskonforme Einbettung von \mathcal{G} in \mathcal{R} .

Abbildung 3.3 zeigt eine spezieskonforme Einbettung von \mathcal{G} in den durch Gen-duplikationen aus \mathcal{S} abgeleiteten Genbaum \mathcal{R} . Der Genbaum \mathcal{R} ist eine Ableitung minimaler Kardinalität aus der unendlichen Menge aller Ableitungen von \mathcal{S} , welche die Inkonsistenz von \mathcal{G} und \mathcal{S} erklären. Einen solchen Genbaum mit minimaler Kardinalität, welchen wir aus \mathcal{S} ableiten und in den \mathcal{G} spezieskonform einbettbar ist, bezeichnen wir als *Reconciled Tree von \mathcal{G} über \mathcal{S}* . Der so beschriebene Reconciled Tree ist von Page (1994a) semiformal definiert worden. Wir werden diese Definition in dem folgenden Abschnitt wiedergeben und in dem darauf folgenden Abschnitt formal beschreiben.

3.2 Die historische Definition

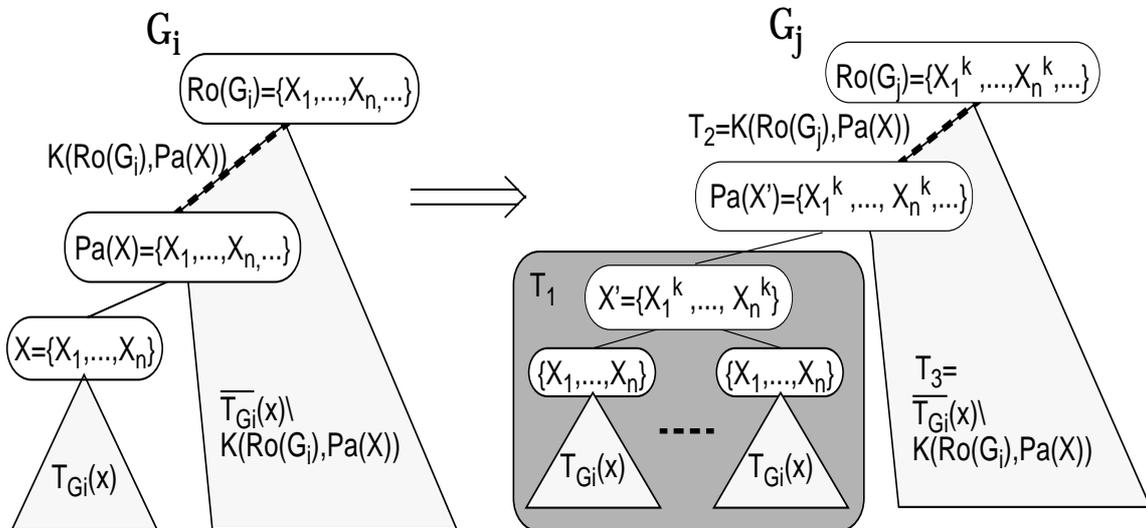
Page hat 1994 den Reconciled Tree durch eine semiformale Definition (Page, 1994a:63) eingeführt. Wir werden seine Definition in den folgenden drei Absätzen zitieren und jeweils deren Bedeutung zeigen.

3.2.1 Ableitung, als n-Clustersystem einer Multimenge

Der Reconciled Tree definiert sich, wie im vorigen Abschnitt gezeigt, über einen Genbaum \mathcal{G} und einen Speziesbaum \mathcal{S} . Page beschreibt diese Bäume in der üblichen Darstellung als n-Clustersysteme ($\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$) derselben Grundmenge M . Für die Darstellung der von \mathcal{S} abgeleiteten Genbäume verwendet Page ein n-Clustersystem über einer Multimenge von M . Die Multimenge definiert Page wie folgt.

Let the set $S = \{x_1, \dots, x_n\}$. A multiset, M_S derived from S is a set $\{x_1^{v_1}, x_2^{v_2}, \dots, x_n^{v_n}\}$ where v_i is the number of times the i th element of S occurs in M . For example, if $M = \{a^1, b^2, c^2, d^2\}$, then $M = \{a, b, b, c, c, d, d\}$. The set S is the base set of M_S and is written S_M .

Abbildungung 3.5 auf Seite 28 zeigt den im vorigen Abschnitt abgeleiteten Genbaum \mathcal{R} (Abbildung 3.3). Die Bezeichnungen ohne alphabetische Zeichen zeigen die Multimengendarstellung von \mathcal{R} . Der Vorteil des Multimengenkonzeptes wird an den Kopien des Duplikationsgens $\{1^3, \dots, 5^3\}$ deutlich. Die Entwicklungen aller Kopien von $\{1^3, \dots, 5^3\}$ tragen dieselben Bezeichnungen. Damit lassen sich auf einfache Weise Genbäume aus S ableiten, ohne für neue Kopien weitere Bezeichnungen einführen zu müssen.



Abbildungung 3.4: Die Ableitung des Genbaumes G_j durch das Duplikationselement a aus dem Speziesbaum S

Wir beschreiben nun, wie Genbäume mit dem Multimengenkonzept aus S induktiv abgeleitet werden. Dazu verwenden wir die binären Relationen \cup und \setminus , welche sich für Multimengen auf natürliche Weise von den gleichbezeichneten binären Relationen auf Mengen ableiten. Als Basis für die Ableitung der Genbäume dient die Ableitung des Genbaumes aus S , welcher keine Genduplikationen enthält. Dies ist das n -Clustersystem S , wobei wir seine Elemente als Multimengen interpretieren.

Sei G_i ein von S abgeleiteter Genbaum und $X \in G_i$. Aus G_i leiten wir einen Genbaum G_j ab, welcher k Kopien des Gens X enthält (siehe Abbildung 3.4). Dazu kopieren wir den Teilbaum $T_{G_i}(X)$ k mal und fügen als deren gemeinsame

Wurzel das Duplikationsgen X' hinzu. X' besteht dabei aus der Vereinigung der Wurzeln aller kopierten Teilbäume, $X' := \bigcup_{i \in \{1, \dots, k\}} \text{Ro}(T_{\mathcal{G}_i}(X))$. Wir erhalten das n -Clustersystem $\mathcal{T}_1 := X' \cup \bigcup_{i \in \{1, \dots, k\}} T_{\mathcal{G}_i}(X)$, welches die k -fache Duplikation an dem Duplikationsgen X' beschreibt. Um einen aus \mathcal{G}_i abgeleiteten Genbaum zu erhalten, fügen wir den Teilbaum $\overline{T_{\mathcal{G}_i}(X)}$ zu \mathcal{T}_1 hinzu. Damit wir durch das Hinzufügen wieder ein n -Clustersystem erhalten, müssen die Elemente der Kette $K_{\mathcal{G}_i}(\text{Ro}(\mathcal{G}_i), \text{Pa}(X))$ die Elemente des Duplikationselementes X' als Teilmenge enthalten. Also fügen wir zu \mathcal{T}_1 die modifizierte Kette $\mathcal{T}_2 := \{V \mid \exists W \in K_{\mathcal{G}_i}(\text{Ro}(\mathcal{G}_i), \text{Pa}(X)) : V = X' \setminus X \cup W\}$ hinzu. Die Elemente $\mathcal{T}_3 := \overline{T_{\mathcal{G}_i}(X)} \setminus K_{\mathcal{G}_i}(\text{Ro}(\mathcal{G}_i), \text{Pa}(X))$ übernehmen wir direkt aus \mathcal{G}_i und erhalten abschließend das neue n -Clustersystem $\mathcal{G}_j := \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3$.

3.2.2 Das m -Clustersystem

Ein so abgeleiteter Genbaum \mathcal{G}_j ist ein n -Clustersystem über einer Multimenge und besitzt die beiden folgenden Eigenschaften.

1. Die Host-Spezies eines Gens $X \in \mathcal{G}_j$ ist dessen Basis S_X (die Basen aller Gene bilden \mathcal{S}).
2. Wenn X dupliziert wurde, besitzen seine Kinder dieselbe Basis, sonst besitzt kein Kind von X dieselbe Basis wie X . Für zwei beliebige unterschiedliche Kinder $a, b \in \text{Ch}(X)$ gilt damit $S_a = S_b = S_X$, wenn X ein Duplikationsgen ist, und $S_a \cap S_b = \emptyset$ sonst.

Mit diesen Eigenschaften eines n -Clustersystemes über einer Multimenge hat Page den m -tree definiert.

An m -tree T is a set of subsets of M satisfying the following conditions:

1. $M \in T$
2. $\emptyset \notin T$
3. $\{i\} \in T$ for all $i \in M$
4. $A \cap B \in \{\emptyset, A, B\}$ for every $A, B \in T$.
5. Let X_i be the cluster of node i . For any two immediate descendants, a and b of node c , $S_{X_a} \cap S_{X_b} = \emptyset$.

In Bedingung 5 sind die direkten Nachfolger als unterschiedlich anzunehmen. Die in der Definition des m -trees verwendeten binären Relationen auf Multimengen sind von Page nicht weiter beschrieben worden und müssen durch den praktischen Hintergrund der Definition verstanden werden. Eine Definition dieser binären Relationen, insbesondere der Relation \in , führt schnell zu einer

unübersichtlichen Formulierung des m -trees. Wir werden später die Multimenge als Menge darstellen, womit der m -tree ein n -Clustersystem mit der obigen Eigenschaft 5 ist. Um mit unserer Bezeichnung n -Clustersystem konsistent zu bleiben, bezeichnen wir den m -tree als m -Clustersystem.

3.2.3 Der Reconciled Tree

Der Reconciled Tree von \mathcal{G} und \mathcal{S} ist somit ein an Kardinalität kleinstes m -Clustersystem, welches von \mathcal{S} ableitbar und in das \mathcal{G} spezieskonform einbettbar ist. Den Speziesbaum bezeichnet Page als Host-Tree T_H , den Genbaum als Associated-Tree T_A und ein n -Clustersystem als n -tree.

Let T_H be an n -tree on the set of hosts H , and let T_A be the set of subsets of H implied by the associated tree ... The reconciled tree T_R for T_H and T_A is the smallest m -tree that

- (1) contains all and only the clusters of T_H and*
- (2) contains T_A as a subtree.*

Die Bedingung (1) beschreibt die Ableitbarkeit des Reconciled Tree von dem Speziesbaum und Bedingung (2) die spezieskonsistente Einbettbarkeit des Genbaumes in den Reconciled Tree.

Das kleinste m -Clustersystem sowie die Beinhaltung von T_A als Teilbaum in T_R ist nicht definiert. Die in der Definition beschriebene Eindeutigkeit des Reconciled Tree ist nicht gezeigt worden. Aufgrund der direkten Ableitung der Definition des Reconciled Tree aus der Genduplikation sind diese Formulierungen jedoch weitgehend verständlich. Für die weitere Arbeit benötigen wir aber eine formale, auf unsere Anforderungen zugeschnittene Definition des Reconciled Tree. Diese werden wir in dem nächsten Abschnitt durch eine Formalisierung der semiformalen Definition von Page definieren.

3.3 Eine formale Definition

Wir werden eine formale Definition der Formulierung des Reconciled Tree von Page geben, welche auf die Erfordernisse dieser Arbeit zugeschnitten ist. Dazu werden wir die von Page verwendeten Begriffe – Multimenge, m -Clustersystem und Reconciled Tree – reformulieren.

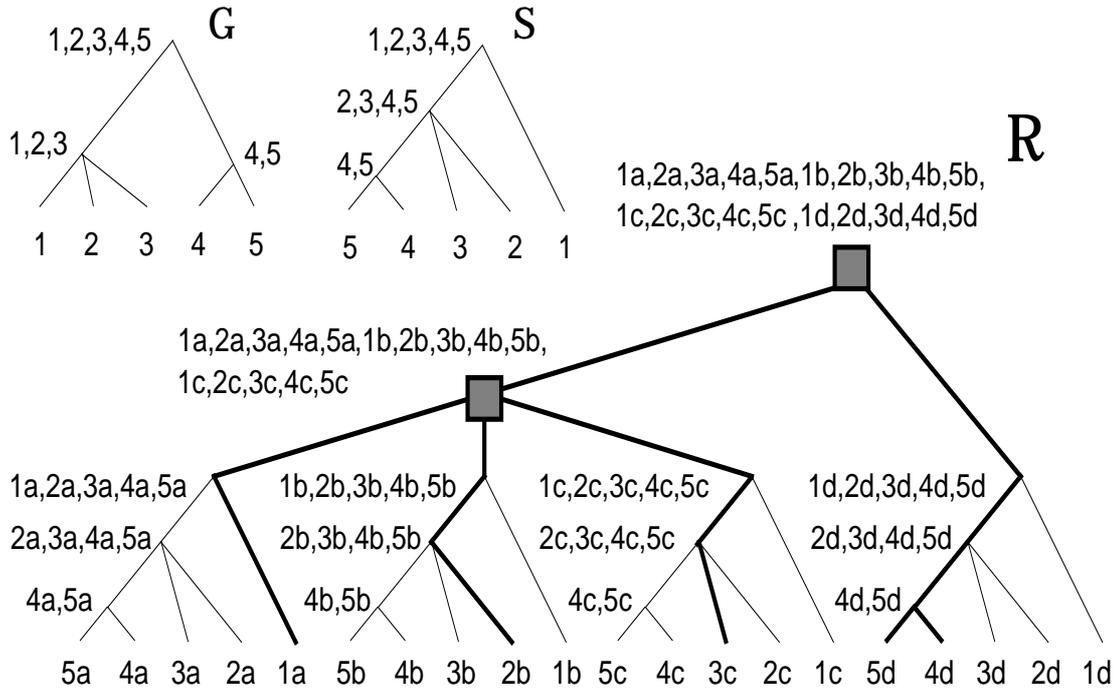


Abbildung 3.5: Der formal definierte Reconciled Tree \mathcal{R} von \mathcal{G} über \mathcal{S}

3.3.1 Reformulierung der Multimenge

Das Verständnis eines Reconciled Tree wird durch die Verwendung der Multimenge erleichtert, indem für Kopien keine neuen Bezeichner eingeführt werden müssen. Aber die nicht eindeutige Bezeichnung der “Elemente” einer Multimenge erschwert den formalen Umgang mit einem Reconciled Tree und auch dessen biologische Analyse. Daher werden wir die Multimenge durch die übliche Mengennotation beschreiben.

Eine Multimenge ist eine “Menge”, in welcher die Vielfachheit von Elementen erlaubt ist. Durch eine eindeutige Indizierung vielfach vorkommender Elemente erhalten wir aus der Multimenge wieder eine “echte” Menge. Z.B. stellen wir die Multimenge $\{a, a, a, b, c, c\}$ durch $\{(a, 1), (a, 3), (a, 7), (b, 2), (b, 7)\}$.

Definition 3.1 (Multimenge)

Sei M eine Menge, dann bezeichnen wir die Menge

$$\text{MULT}(M) := \wp(M \times \mathbb{N})$$

als die Menge aller Multimengen über der Menge M . Ein Element aus $M_{\text{ult}} \in \text{MULT}(M)$ bezeichnen wir als Multimenge. Wir bezeichnen $v(k) := |\{(x, i) \in M_{\text{ult}} \mid x = k\}|$ als die Vielfachheit des Elementes k in M .

Die informale “Mengenklammer-Notation” der Multimenge über einer Menge M von Page beschreibt genau deren Vielfachheit. Ein Beispiel: Sei $M := \{1, 2, 3, 4\}$ und $v(1) = 1$, $v(2) = 2$, $v(3) = 1$, $v(4) = 0$. In der informalen Notation von Page entspricht dies $\{1, 2, 2, 3\}$ und in unserer Multimengennotation beispielsweise $\{(1, 1), (2, 1), (2, 2), (3, 1)\}$.

Die Abbildung 3.5 zeigt den in Abbildung 3.3 abgeleiteten Reconciled Tree \mathcal{R} in der eben definierten Multimengennotation. Der Übersicht wegen ist die Beschreibung der Knoten umformuliert. So ist z.B. die Bezeichnung $1a, 2a, 3a, 4a, \dots, 1c, 2c, 3c, 4c$ die vereinfachte Darstellung der Bezeichnung $(1, a), (2, a), (3, a), (4, a), \dots, (1, c), (2, c), (3, c), (4, c)$. Hierbei beschreiben die alphabetischen Zeichen die natürlichen Zahlen von 1 bis 4 mit $a = 1$, $b = 2$, $c = 3$ und $d = 4$.

Wir übertragen die Formulierung der Basis einer Multimenge von Page auf die von uns eingeführte Multimengennotation.

Definition 3.2 (Basis einer Multimenge)

Sei N eine Menge, $M \in \text{MULT}(N)$ und $\mathcal{M} \subseteq \wp(M)$.

1. Wir bezeichnen $B(M) := \{x \in N \mid (x, i) \in M\}$ als die Basis von M .
2. Wir bezeichnen $B(\mathcal{M}) := \bigcup_{Y \in \mathcal{M}} B(Y)$ als die Basis von \mathcal{M} .
3. Die Abbildung $B_{\mathcal{M}} : \mathcal{M} \rightarrow N$ mit $B_{\mathcal{M}}(Y) \mapsto B(Y)$ bezeichnen wir als die Abbildung der Basis B bezogen auf \mathcal{M} .

Abschließend zeigen wir einige im weiteren verwendete Eigenschaften der Basis einer Multimenge.

Lemma 3.1

Sei M eine Menge, $\mathcal{M} \in \text{MULT}(M)$ und $X, Y \in \mathcal{M}$. Dann gilt

$$Y \subseteq X \Rightarrow B(Y) \subseteq B(X).$$

(Es gilt weder $Y \subset X \Rightarrow B(Y) \subset B(X)$ noch $B(Y) \subseteq B(X) \Rightarrow Y \subseteq X$.)

Beweis. Sei $y \in B(Y)$, dann existiert ein $i \in N$ mit $(y, i) \in Y$. Aus $Y \subseteq X$ folgt $(y, i) \in X$ und somit $y \in B(X)$. \square

Verw. von: ; Verw. in: Lem. 5.2, Theo. 4.6.

Lemma 3.2

Sei M eine Menge, $\mathcal{M} \in \text{MULT}(M)$ und $X, Y \in \mathcal{M}$, dann gilt

$$B(X) \setminus B(Y) \subseteq B(X \setminus Y).$$

Beweis. Sei $e \in B(X) \setminus B(Y)$, dann ist $e \in B(X)$ und $e \notin B(Y)$. Damit existiert ein $i \in \mathbb{N}$ mit $(e, i) \in X$ und $(e, i) \notin Y$. Hieraus folgt $(e, i) \in X \setminus Y$ und weiter $e \in B(X \setminus Y)$. \square

Lemma 3.3

Sei M eine Menge und $X, Y \in \text{Mult}(M)$, dann gilt

$$B(X) \setminus B(Y) \neq B(X \setminus Y) \Leftrightarrow \exists (e, i) \in X \setminus Y \wedge \exists (e, j) \in Y.$$

Beweis.

“ \Leftarrow ”: Es gelte $\exists (e, i) \in X \setminus Y \wedge \exists (e, j) \in Y$. Damit ist $e \in B(X \setminus Y)$, also gilt auch $e \in B(X)$. Da $(e, j) \in Y$ gilt, folgt $e \in B(Y)$. Aus $e \in B(X)$ und $e \in B(Y)$ folgt $e \notin B(X) \setminus B(Y)$. Somit ist $B(X \setminus Y) \neq B(X) \setminus B(Y)$.

“ \Rightarrow ”: Es gelte $B(X) \setminus B(Y) \neq B(X \setminus Y)$. Mit Lemma 3.2 folgt hieraus $B(X) \setminus B(Y) \subset B(X \setminus Y)$, womit ein $e \in B(X \setminus Y)$ mit $e \notin B(X) \setminus B(Y)$ existiert.

Aus $e \in B(X \setminus Y)$ folgt, es existiert ein $i \in \mathbb{N}$ mit $(e, i) \in X \setminus Y$.

Aus $e \notin B(X) \setminus B(Y)$ folgt entweder $e \notin B(X)$ oder $e \in B(X)$ und $e \in B(Y)$. Da $e \in B(X \setminus Y)$ ist, ist $e \in B(X)$, womit $e \in B(X)$ und $e \in B(Y)$ folgt. Also existiert ein $j \in \mathbb{N}$ mit $(e, j) \in Y$. \square

Verw. von: Lem. 3.2; Verw. in: Lem. 4.9, 4.13, Theo. 4.8.

3.3.2 Reformulierung des m-Clustersystemes

Wir reformulieren das m-Clustersystem von Page, welches auf Seite 26 beschrieben ist. Die Eigenschaften 1 bis 4 des m-Clustersystemes von Page beschreiben ein n-Clustersystem (Def. 2.4, Seite 17) über einer Multimenge. Damit ist ein m-Clustersystem ein n-Clustersystem auf Multimengen mit der Eigenschaft 5 des m-Clustersystemes von Page.

Für einen abgeleiteten Genbaum \mathcal{G}_j und $X \in \mathcal{G}_j$ gilt nach Bedingung 5 des m-Clustersystemes von Page

$$\forall Z_i, Z_j \in \text{Ch}(X) : Z_i \neq Z_j \Rightarrow B(Z_i) = B(Z_j) = B(X) \vee B(Z_i) \cap B(Z_j) = \emptyset.$$

Wir verwenden die hierzu äquivalente Aussage

$$\forall Z_i, Z_j \in \text{Ch}(X) : Z_i \neq Z_j \wedge B(Z_i) \cap B(Z_j) \neq \emptyset \Rightarrow B(Z_i) = B(Z_j) = B(X).$$

Mit dieser äquivalenten Bedingung schränken wir die n -Clustersysteme auf Multimengen ein und erhalten in Analogie zu dem m -Clustersystem von Page die folgende Reformulierung des m -Clustersystemes.

Definition 3.3 (m-Clustersystem)

Sei \mathcal{N} ein n -Clustersystem über einer Multimenge M . Wir bezeichnen \mathcal{N} als m -Clustersystem über M , wenn:

$$\forall X \in \mathcal{N} : \forall Z_i, Z_j \in \text{Ch}(X) : \\ Z_i \neq Z_j \wedge B(Z_i) \cap B(Z_j) \neq \emptyset \Rightarrow B(Z_i) = B(Z_j) = B(X).$$

Die Mengen der Basen von zwei Teilbäumen, gewurzelt an Kopien desselben Duplikationsgens einer Ableitung von einem Speziesbaum, unterscheiden sich nicht. In einem m -Clustersystem ist dies jedoch möglich. Somit kann ein m -Clustersystem noch weiter eingeschränkt werden, bis nur noch Ableitungen aus einem Speziesbaum beschrieben werden können. Jedoch ist diese Einschränkung in dem nächsten Abschnitt schon durch Anbindung des m -Clustersystemes an einen Speziesbaum (von dem dieses abgeleitet wird) gegeben.

3.3.3 Reformulierung des Reconciled Tree

Der Reconciled Tree \mathcal{R} von einem Genbaum \mathcal{G} über einem Speziesbaum \mathcal{S} ist ein von \mathcal{S} abgeleitetes m -Clustersystem minimaler Kardinalität, in welchen \mathcal{G} spezieskonform einbettbar ist. Ein m -Clustersystem ist genau dann von \mathcal{S} ableitbar, wenn $B(\mathcal{R}) = \mathcal{S}$ gilt. Da wir später häufig Ketten verwenden, definieren wir diese Eigenschaft direkt auf diesen.

Definition 3.4 (Ableitbarkeit)

Sei M eine Menge, \mathcal{S} ein n -Clustersystem über M und \mathcal{R} ein m -Clustersystem über einer Multimenge von M . Wenn

$$B(K_{\mathcal{R}}(X, Y)) = K_{\mathcal{S}}(B(X), B(Y)) \text{ für alle } X, Y \in \mathcal{R} \text{ mit } Y \subseteq X$$

gilt, ist \mathcal{R} von \mathcal{S} ableitbar.

Eine Einbettung von \mathcal{G} in \mathcal{R} bedeutet, daß ein Monomorphismus bzgl. der kleinsten oberen Schranke \sqcup von \mathcal{G} nach \mathcal{R} existiert. Daher beschreiben wir diese Einbettung durch einen Einbettungsmonomorphismus von \mathcal{G} nach \mathcal{S} . Die Funktion $\varphi \in \text{Emon}(\mathcal{G}, \mathcal{R})$ ist eine solche Einbettung. Damit \mathcal{G} spezieskonform zu \mathcal{R} ist, muß die Spezies $B(\varphi(X))$ der Einbettung $\varphi(X) \in \mathcal{R}$ eines Gens $X \in \mathcal{G}$ auch das Gen X beinhalten können. Dies ist dann der Fall, wenn $X \subseteq B(\varphi(X))$ gilt.

Definition 3.5 (Spezieskonforme Einbettung)

Sei \mathcal{G} ein n -Clustersystem über einer Menge M , und \mathcal{R} ein m -Clustersystem über einer Multimenge von M . Wenn ein Einbettungsmonomorphismus

$$\varphi \in \text{Emon}(\mathcal{G}, \mathcal{R}) \text{ mit } X \subseteq B(\varphi(X)) \text{ für alle } X \in \mathcal{G} \text{ existiert,}$$

bezeichnen wir φ als spezieskonforme Einbettung von \mathcal{G} in \mathcal{R} .

Lemma 3.4

Sei M eine Menge, $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$, \mathcal{R} ein von \mathcal{S} ableitbares m -Clustersystem und φ eine spezieskonforme Einbettung von \mathcal{G} in \mathcal{R} , dann gilt

$$\bigsqcup_{\mathcal{S}}(X) \subseteq B(\varphi(X)) \text{ für alle } X \in \mathcal{G}.$$

Beweis. Sei $X \in \mathcal{G}$. Aus der Ableitbarkeit von \mathcal{R} aus \mathcal{S} folgt $B(\varphi(X)) \in \mathcal{S}$. Da φ spezieskonform ist, gilt $X \subseteq B(\varphi(X))$. Das kleinste Element in \mathcal{S} , welches X enthält, ist $\bigsqcup_{\mathcal{S}}(X)$. Daraus folgt mit der Baumordnung von \mathcal{S} , daß $\bigsqcup_{\mathcal{S}}(X) \subseteq B(\varphi(X))$ gilt. \square

Verwendung von: ; Verwendung in: Theorem 4.4

Definition 3.6 (Reconciled Tree)

Sei M eine Menge und $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$. Ein an Kardinalität minimales m -Clustersystem \mathcal{R} über einer Multimenge von M , welches den Bedingungen

1. \mathcal{R} ist von \mathcal{S} ableitbar
2. es existiert eine spezieskonforme Einbettung von \mathcal{G} in \mathcal{R}

genügt, bezeichnen wir als Reconciled Tree von \mathcal{G} über \mathcal{S} . Die Menge aller Reconciled Trees von \mathcal{G} über \mathcal{S} bezeichnen wir durch $\text{RC}(\mathcal{G}, \mathcal{S})$.

Ein Duplikationsgen X in einem Reconciled Tree besitzt die Eigenschaft, daß alle seine Kinder als Duplikate in derselben Spezies beinhaltet sind wie X . Damit gilt $B(Z_i) = B(Z_j) = B(X)$ für zwei beliebige ungleiche Kinder $Z_i, Z_j \in \text{FCh}(X)$ von X . In einem m -Clustersystem folgt $B(Z_i) = B(Z_j) = B(X)$ aus $B(Z_i) \cap B(Z_j) \neq \emptyset$. Damit ist X ein Duplikationsgen, wenn $B(Z_i) \cap B(Z_j) \neq \emptyset$ gilt.

Definition 3.7 (Duplikationselement)

Sei M eine Menge, $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$, $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$ und $X'' \in \mathcal{R}$.

$$\text{Existieren ungleiche } Z_i, Z_j \in \text{Ch}(X'') \text{ mit } B(Z_i) \cap B(Z_j) \neq \emptyset,$$

dann bezeichnen wir X als Duplikationselement. In \mathcal{R} bezeichnen wir die Menge aller Duplikationselemente durch $\text{Dup}_{\mathcal{R}}$ und das Komplement von $\text{Dup}_{\mathcal{R}}$ in \mathcal{R} als $\overline{\text{Dup}}_{\mathcal{R}}$. Ist \mathcal{R} offensichtlich, verwenden wir jeweils die Bezeichnungen Dup und $\overline{\text{Dup}}$.

Für den Reconciled Tree \mathcal{R} in Abbildung 3.5 (Seite 28) gilt beispielsweise

$$\begin{aligned} \text{Dup}_{\mathcal{R}} = & \\ & \{ \{ (1, a), (2, a), (3, a), (4, a), (5, a), \dots, (1, c), (2, c), (3, c), (4, c), (5, c) \}, \\ & \{ (1, a), (2, a), (3, a), (4, a), (5, a), \dots, (1, d), (2, d), (3, d), (4, d), (5, d) \} \}. \end{aligned}$$

Kapitel 4

Strukturaussagen zum Reconciled Tree

Der Reconciled Tree \mathcal{R} eines Genbaumes \mathcal{G} über einem Speziesbaum \mathcal{S} ist biologisch einfach zu interpretieren, da dieser direkt aus der Evolution abgeleitet worden ist. Jedoch wissen wir wenig über die Struktur des Reconciled Tree \mathcal{R} und haben damit keine konkrete Vorstellung davon, wie der Genbaum \mathcal{G} in den Reconciled Tree \mathcal{R} eingebettet ist, welche Duplikationselemente \mathcal{R} enthält, welche Topologie \mathcal{R} besitzt oder ob \mathcal{R} eindeutig in $\text{RC}(\mathcal{G}, \mathcal{S})$ bestimmt ist. Die hierzu wesentlichen Aussagen werden wir in diesem Kapitel zeigen. Diese sind wichtig, da wir hierdurch den Reconciled Tree in Kapitel 5 konstruktiv (direkt) beschreiben und damit schnell berechnen können.

Um die Aussagen dieses Kapitels zu zeigen, konstruieren wir uns zunächst einige Hilfsmittel. Dies sind Operationen, welche wir in Abschnitt 4.1 definieren, mit denen wir m -Clustersysteme in andere m -Clustersysteme umwandeln können. Unter bestimmten Voraussetzungen können wir mit diesen Operationen einen Reconciled Tree (dieser ist ein m -Clustersystem) in einen anderen Reconciled Tree überführen, welches wir in Abschnitt 4.2 zeigen.

Mit diesen Operationen lassen sich dann wesentliche Aussagen zu dem Reconciled Tree nach dem gleichen Konzept (durch Widerspruch) zeigen. Bei diesem Konzept gehen wir davon aus, daß ein Reconciled Tree eine zu zeigende Eigenschaft E nicht besitzt. Mit den Operationen auf m -Clustersystemen konstruieren wir hieraus einen Reconciled Tree, der eine geringere Kardinalität als der ursprüngliche besitzt. Da jeder Reconciled Tree zu dem gleichen Gen- und Speziesbaum dieselbe Kardinalität besitzt, existiert somit kein Reconciled Tree, welcher die Eigenschaft E nicht besitzt.

In Abschnitt 4.3 zeigen wir mit den Operationen auf m -Clustersystemen das

LCA-Theorem, welches eine wesentliche Aussage über die Einbettung des Genbaumes \mathcal{G} in den Reconciled Tree \mathcal{R} macht.

Mit den Operationen auf m-Clustern und dem LCA-Theorem zeigen wir dann in Abschnitt 4.4 die bis auf Symmetrie eindeutige Topologie des Reconciled Tree \mathcal{R} . Diese setzt sich aus Subtopologien von \mathcal{S} zusammen, welche durch die Topologie von \mathcal{G} bestimmt sind.

Um die Topologie zu zeigen, verwenden wir die zentrale Aussage des Duplikations-Theorems, mit welchem sich alle Duplikationselemente in \mathcal{R} nur durch Elemente in \mathcal{G} und \mathcal{S} beschreiben lassen. Mit dem Duplikations-Theorem lassen sich somit Genduplikationen auch ohne die direkte Konstruktion des Reconciled Tree bestimmen, womit sich diese schneller berechnen lassen als der vollständige Reconciled Tree.

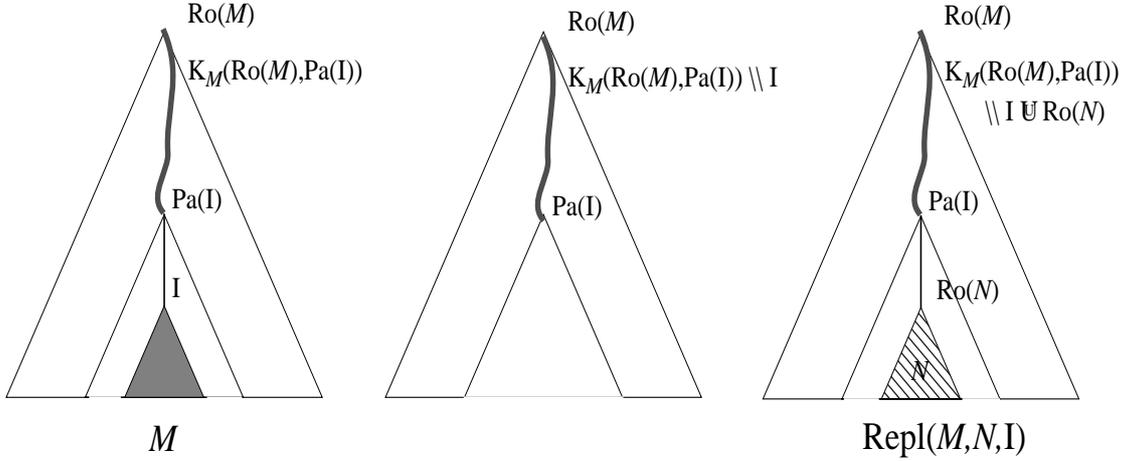
Allgemein werden wir Variablen ohne Apostroph für Elemente in einem Genbaum, mit einem Apostroph für Elemente in einem Speziesbaum und mit zwei Apostrophen für Elemente eines Reconciled Tree verwenden (z.B. $X \in \mathcal{G}$, $X' \in \mathcal{S}$, $X'' \in \mathcal{R}$).

4.1 Operationen auf m-Clustern

In diesem Kapitel führen wir Operationen auf m-Clustern ein und zeigen, daß diese wohldefiniert sind. Die Operation Repl ($\text{Repl} \hat{=} \text{replace}$) in Abschnitt 4.1.1 ersetzt einen Teilbaum eines m-Clustersystems durch ein anderes m-Clustersystem, die Operation Merge in Abschnitt 4.1.2 fügt eine Menge von m-Clustern zu einem neuen m-Clustern zusammen, und die Operation \times in Abschnitt 4.1.3 wandelt ein n-Clustern in ein topologisch äquivalentes m-Clustern um.

4.1.1 Die Operation: Replace

Wie in Abbildung 4.1 grafisch gezeigt, ersetzt die Operation $\text{Repl}(\mathcal{M}, \mathcal{N}, I)$ in dem m-Clustern \mathcal{M} den Teilbaum $T_{\mathcal{M}}(I)$ durch das m-Clustern \mathcal{N} . Wir definieren die Ersetzung derart, daß $\mathcal{L} := \text{Repl}(\mathcal{M}, \mathcal{N}, I)$ wieder ein m-Clustern ist. In dem nächsten Abschnitt 4.1.1.1 formulieren wir zuerst die Ersetzung und zeigen in dem darauf folgenden Abschnitt 4.1.1.2 die Wohldefiniertheit der Ersetzung.

Abbildung 4.1: Die Ersetzung $\text{Repl}(\mathcal{M}, \mathcal{N}, I)$

4.1.1.1 Formulierung der Ersetzung

Die Ersetzung \mathcal{L} kann in das “Entfernen” des Teilbaumes $T_{\mathcal{M}}(I)$ und das “Hinzufügen” des m -Clustersystemes \mathcal{N} untergliedert werden.

Entfernen von $T_{\mathcal{M}}(I)$:

Um $T_{\mathcal{M}}$ aus \mathcal{M} zu “entfernen”, reicht es nicht aus, \mathcal{M} um $T_{\mathcal{M}}$ zu reduzieren. Da \mathcal{M} ein n -Clustersystem ist, enthalten alle Mengen der Kette $\mathcal{K} := K_{\mathcal{M}}(\text{Ro}(\mathcal{M}), \text{Pa}(I))$ die Wurzel I von $T_{\mathcal{M}}(I)$ als echte Teilmenge. Wir reduzieren also zusätzlich jede Menge aus $K_{\mathcal{M}}(\text{Ro}(\mathcal{M}), \text{Pa}(I))$ um die Menge I .

Um allgemein jede Menge eines gegebenen Mengensystemes durch eine beliebige Menge zu reduzieren, führen wir die binäre Mengenoperation \parallel auf Mengensystemen als Erweiterung der Mengenoperation \setminus auf Mengen ein.

Definition 4.1 (\parallel)

Sei \mathcal{O} ein Mengensystem und M eine Menge, dann ist

$$\mathcal{O} \parallel M := \{X \mid \exists Z \in \mathcal{O} : X = Z \setminus M\}$$

die Mengensubstitution auf Mengensystemen.

Damit ergibt sich das Mengensystem $\mathcal{M}_{\text{red}} := \mathcal{M} \setminus (\mathcal{K} \cup T_{\mathcal{M}}(I)) \cup (\mathcal{K} \parallel I)$ (siehe mittlere Skizze in Abbildung 4.1) als die *Entfernung* von $T_{\mathcal{M}}(I)$ in \mathcal{M} . \mathcal{M}_{red} kann auch durch $\mathcal{M} \parallel I \setminus \{\emptyset\}$ ausgedrückt werden.

Die Entfernung selbst ist im allgemeinen nicht wohldefiniert. Angenommen, $\text{Pa}(I)$ besitzt in \mathcal{M} mindestens drei Kinder, so ist \mathcal{M}_{red} offensichtlich wieder ein m -Clustersystem. Ansonsten ist \mathcal{M}_{red} kein m -Clustersystem, da die Bedingung

des m -Clustersystemes, daß jeder innere Knoten mindestens zwei Kinder besitzt, verletzt ist. Also können wir nicht direkt die Operation Repl durch zwei primitivere Operationen wie “Entfernen” und “Hinzufügen” ersetzen, welche wohldefiniert sind.

Hinzufügen von \mathcal{N} :

Wir fügen das m -Clustersystem \mathcal{N} zu dem Mengensystem \mathcal{M}_{red} hinzu. Hierzu hat \mathcal{N} den zwei folgenden Bedingungen zu genügen, damit wir wieder ein m -Clustersystem erhalten.

1. Die Eigenschaft, ein n -Clustersystem aus der Vereinigung von \mathcal{N} und \mathcal{M}_{red} zu bilden, darf durch \mathcal{N} nicht “gestört” werden: Das reduzierte Mengensystem \mathcal{M}_{red} ist bis auf die Möglichkeit, daß das Element $\text{Pa}_{\mathcal{M}}(\text{Ro}(\text{T}_{\mathcal{M}}(I)))$ nur ein Kind besitzt, ein n -Clustersystem und \mathcal{N} ist nach Voraussetzung ein n -Clustersystem. Also sind diese beiden Mengensysteme Bäume. Damit die Baumeigenschaft der Vereinigung dieser beiden Mengensysteme nicht verletzt wird, formulieren wir für \mathcal{N} die Voraussetzung $\mathcal{M}_{\text{red}} \cap \mathcal{N} = \emptyset$. Wir verwenden im folgenden hierzu die einfachere Voraussetzung $\text{Ro}(\mathcal{M}_{\text{red}}) \cap \text{Ro}(\mathcal{N}) = \emptyset$, welche $\mathcal{M}_{\text{red}} \cap \mathcal{N} = \emptyset$ impliziert. (Da $\text{Ro}(\mathcal{M}_{\text{red}}) \subseteq \text{Ro}(\mathcal{M})$ gilt, folgt $\text{Ro}(\mathcal{M}) \cap \text{Ro}(\mathcal{N}) = \emptyset \Rightarrow \text{Ro}(\mathcal{M}_{\text{red}}) \cap \text{Ro}(\mathcal{N}) = \emptyset$. $\text{Ro}(\mathcal{M}_{\text{red}})$ und $\text{Ro}(\mathcal{N})$ sind jeweils die Maxima von \mathcal{M}_{red} und \mathcal{N} . Also gilt $\text{Ro}(\mathcal{M}_{\text{red}}) \cap \text{Ro}(\mathcal{N}) = \emptyset \Rightarrow \mathcal{M}_{\text{red}} \cap \mathcal{N} = \emptyset$. Somit gilt $\text{Ro}(\mathcal{M}_{\text{red}}) \subseteq \text{Ro}(\mathcal{M}) \Rightarrow \mathcal{M}_{\text{red}} \cap \mathcal{N} = \emptyset$.)
2. Die Eigenschaft, ein m -Clustersystem aus der Vereinigung von \mathcal{N} und \mathcal{M}_{red} zu bilden, darf durch \mathcal{N} nicht “gestört” werden: Dazu “erhalten” wir die Duplikationselemente von $\overline{\text{T}_{\mathcal{M}}(I)}$ in der Vereinigung von \mathcal{N} mit dem reduzierten Mengensystem \mathcal{M}_{red} . \mathcal{M}_{red} ist bis auf die Möglichkeit, daß das Element $\text{Pa}_{\mathcal{M}}(\text{Ro}(\text{T}_{\mathcal{M}}(I)))$ nur ein Kind besitzt, ein m -Clustersystem. Die Eigenschaft, ob $\text{Pa}(I)$ ein Duplikationselement in \mathcal{M} ist oder nicht, soll auf das Element $\text{Pa}_{\mathcal{M}}(\text{Ro}(\text{T}_{\mathcal{M}}(I)))$ in der Vereinigung von \mathcal{M}_{red} und \mathcal{N} übertragen werden. Dies sichern wir durch die Voraussetzung $\text{B}(\text{Ro}(\mathcal{N})) = \text{B}(I)$ an \mathcal{N} .

Wir fügen nun ein m -Clustersystem \mathcal{N} , welches den eben genannten Bedingungen entspricht, dem Mengensystem \mathcal{M}_{red} durch Vereinigung mit diesem hinzu. Damit die Vereinigung von \mathcal{M}_{red} mit \mathcal{N} ein m -Clustersystem ergibt, muß jedes Element der Kette $\mathcal{K}(\text{Ro}(\mathcal{M}_{\text{red}}), \text{Pa}_{\mathcal{M}}(I) \setminus I)$ in \mathcal{M}_{red} mit $\text{Ro}(\mathcal{N})$ vereinigt werden. Wir führen hierzu allgemein die binäre Mengenoperation \uplus auf Mengensystemen als Erweiterung der Mengenoperation \cup auf Mengen ein.

Definition 4.2 (\uplus)

Sei \mathcal{O} ein Mengensystem und M eine Menge, dann ist

$$\mathcal{O} \uplus M := \{X \mid \exists Z \in \mathcal{O} : X = Z \cup M\}$$

die Mengenvereinigung auf Mengensystemen.

Somit beschreiben wir das Hinzufügen (rechte Skizze in Abbildung 4.1) durch

$$\begin{aligned} & \mathcal{M}_{\text{red}} \cup ((\mathcal{K} \setminus I) \uplus \text{Ro}(\mathcal{N})) \cup \mathcal{N} \\ &= \mathcal{M} \setminus (\mathcal{K} \cup T_{\mathcal{M}}(I)) \cup ((\mathcal{K} \setminus I) \uplus \text{Ro}(\mathcal{N})) \cup \mathcal{N} \\ &= \overline{T_{\mathcal{M}}(I)} \setminus K \cup \mathcal{N} \cup ((\mathcal{K} \setminus I) \uplus \text{Ro}(\mathcal{N})) \end{aligned}$$

Also wird für die Ersetzung die Kette \mathcal{K} in die Kette $(\mathcal{K} \setminus I) \uplus \text{Ro}(\mathcal{N})$ umgewandelt, die Elemente $\overline{T_{\mathcal{M}}(I)} \setminus \mathcal{K}$ beibehalten und \mathcal{N} hinzugefügt. Die Ersetzung der Kette und die Beibehaltung der Elemente läßt sich durch die folgende Abbildung ausdrücken.

Definition 4.3 ($\Upsilon_{\mathcal{M},\mathcal{N},I}$)

Sei \mathcal{M} ein m -Clustersystem, $I \in \mathcal{M}$ und \mathcal{N} ein m -Clustersystem mit $\text{Ro}(\mathcal{N}) \cap \text{Ro}(\mathcal{M}) = \emptyset$, sowie $B(\text{Ro}(\mathcal{N})) = B(I)$. Wir definieren die Funktion $\Upsilon_{\mathcal{M},\mathcal{N},I} : \overline{T_{\mathcal{M}}(I)} \rightarrow \overline{T_{\mathcal{M}}(I)} \setminus K_{\mathcal{M}}(\text{Ro}(\mathcal{M}), \text{Pa}(I)) \cup ((K_{\mathcal{M}}(\text{Ro}(\mathcal{M}), \text{Pa}(I)) \setminus I) \uplus \text{Ro}(\mathcal{N}))$ durch

$$\Upsilon_{\mathcal{M},\mathcal{N},I}(X) := \begin{cases} X \setminus I \cup \text{Ro}(\mathcal{N}) & X \in K_{\mathcal{M}}(\text{Ro}(\mathcal{M}), \text{Pa}(I)) \\ X & \text{sonst.} \end{cases}$$

Die Abbildung von $\Upsilon_{\mathcal{M},\mathcal{N},I}$ auf ihrem Urbildbereich, vereinigt mit \mathcal{N} , ergibt nun die Ersetzung $\text{Repl}(\mathcal{M}, \mathcal{N}, I)$.

Definition 4.4 (Repl)

Wir bezeichnen

$$\text{Repl}(\mathcal{M}, \mathcal{N}, I) := \Upsilon_{\mathcal{M},\mathcal{N},I}(\mathcal{M} \setminus T_{\mathcal{M}}(I)) \cup \mathcal{N}$$

als die Ersetzung von \mathcal{N} an Element I in \mathcal{M} . Abkürzend verwenden wir, wenn die Mengen offensichtlich sind, den Begriff der Ersetzung.

4.1.1.2 Wohldefiniertheit der Ersetzung

Wir zeigen nun, daß die Ersetzung wohldefiniert ist. Hierzu formulieren wir für diesen Abschnitt die folgenden Festlegungen: \mathcal{M} ist ein m -Clustersystem, $I \in \mathcal{M}$ und \mathcal{N} ein m -Clustersystem mit $\text{Ro}(\mathcal{N}) \cap \text{Ro}(\mathcal{M}) = \emptyset$ sowie $B(\text{Ro}(\mathcal{N})) = B(I)$ und $\mathcal{K} := K_{\mathcal{M}}(\text{Ro}(\mathcal{M}), \text{Pa}(I))$.

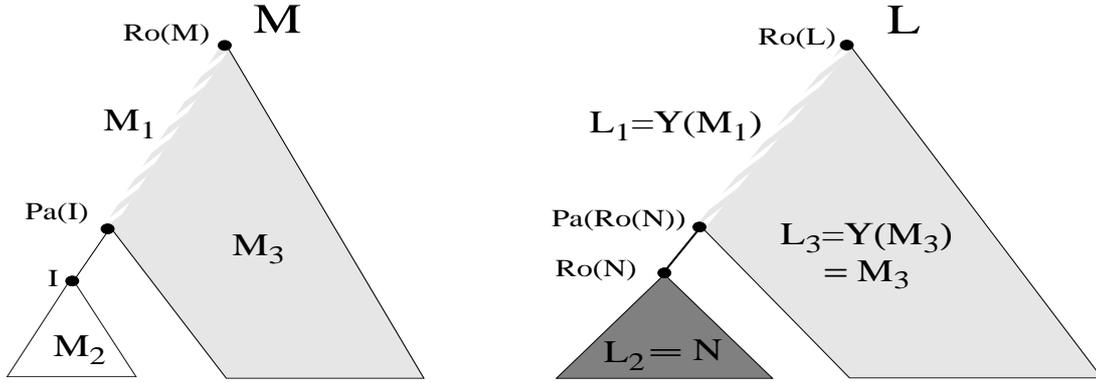


Abbildung 4.2: Die Partitionierung von \mathcal{M} und der Ersetzung $\mathcal{L} := \text{Repl}(\mathcal{M}, \mathcal{N}, I)$

Um den Beweis zu vereinfachen, erweitern wir die Abbildung $\Upsilon_{\mathcal{M}, \mathcal{N}, I}$ um das Element \emptyset durch die Funktion Υ .

Definition 4.5

Wir definieren die Funktion

$$\Upsilon : \overline{\mathbb{T}_{\mathcal{M}}}(I) \cup \{\emptyset\} \rightarrow \overline{\mathbb{T}_{\mathcal{M}}}(I) \setminus \mathcal{K} \cup ((\mathcal{K} \setminus I) \uplus \text{Ro}(\mathcal{N})) \cup \{\emptyset\} \text{ durch}$$

$$\Upsilon|_{\overline{\mathbb{T}_{\mathcal{M}}}(I)} := \Upsilon_{\mathcal{M}, \mathcal{N}, I} \text{ und } \Upsilon(\emptyset) := \emptyset.$$

Wir partitionieren das Mengensystem \mathcal{M} und leiten daraus mit Υ die in Abbildung 4.2 gezeigte Partition der Ersetzung $\mathcal{L} := \text{Repl}(\mathcal{M}, \mathcal{N}, I)$ ab.

Partition auf \mathcal{M}	Partition auf \mathcal{L}
$\mathcal{M}_1 := \mathcal{K}$	$\mathcal{L}_1 := \Upsilon(\mathcal{M}_1)$
$\mathcal{M}_2 := \mathbb{T}_{\mathcal{M}}(I)$	$\mathcal{L}_2 := \mathcal{N}$
$\mathcal{M}_3 := \overline{\mathbb{T}_{\mathcal{M}}}(I) \setminus \mathcal{K}$	$\mathcal{L}_3 := \Upsilon(\mathcal{M}_3)$

Außerdem erweitern wir die Mengen \mathcal{M}_3 und \mathcal{L}_3 um das Element $\{\emptyset\}$ mit $\mathcal{M}'_3 := \mathcal{M}_3 \cup \{\emptyset\}$ und $\mathcal{L}'_3 := \mathcal{L}_3 \cup \{\emptyset\}$. Die Ordnungen der Mengen $\mathcal{M}_1 \cup \mathcal{M}'_3$ und $\mathcal{L}_1 \cup \mathcal{L}'_3$ entsprechen sich, da Υ ein Ordnungsisomorphismus ist, wie wir im folgenden zeigen.

Lemma 4.1

Υ ist eine bijektive Abbildung mit

$$\Upsilon^{-1}(Y) = \begin{cases} Y \setminus \text{Ro}(\mathcal{N}) \cup I & Y \in \mathcal{L}_1 \\ Y & \text{sonst.} \end{cases}$$

Beweis.

Offensichtlich gilt $\Upsilon^{-1}(\Upsilon(X)) = X$ und $\Upsilon(\Upsilon^{-1}(Y)) = Y$, für $X \in \mathcal{M}'_3$ und $Y \in \mathcal{L}'_3$.
Für $X \in \mathcal{M}_1$ gilt:

$$\begin{aligned}\Upsilon^{-1}(\Upsilon(X)) &= \Upsilon^{-1}(X \setminus I \cup \text{Ro}(\mathcal{N})) \\ &= (X \setminus I \cup \text{Ro}(\mathcal{N})) \setminus \text{Ro}(\mathcal{N}) \cup I \\ &= (X \setminus I) \cup I \\ &= X \quad (\text{da } I \subset X \text{ gilt})\end{aligned}$$

Für $Y \in \mathcal{L}_1$ gilt:

$$\begin{aligned}\Upsilon(\Upsilon^{-1}(Y)) &= \Upsilon(Y \setminus \text{Ro}(\mathcal{N}) \cup I) \\ &= (Y \setminus \text{Ro}(\mathcal{N}) \cup I) \setminus I \cup \text{Ro}(\mathcal{N}) \\ &= Y \setminus (\text{Ro}(\mathcal{N}) \cup \text{Ro}(\mathcal{N})) \\ &= Y \cup \text{Ro}(\mathcal{N}) \\ &= Y \quad (\text{da } \text{Ro}(\mathcal{N}) \subset Y \text{ gilt})\end{aligned}$$

□

Verw. von: ; *Verw. in:* Lem. 4.2, 4.3, 4.1, Kor. 4.2.

Lemma 4.2

Υ ist ein Ordnungsisomorphismus.

Beweis. Seien $A, B \in \mathcal{M}_1 \cup \mathcal{M}'_3$ mit $A \subseteq B$. Wir betrachten die folgenden Fälle:

1. $A, B \in \mathcal{M}'_3$: Offensichtlich ist $\Upsilon|_{\mathcal{M}'_3}$ ein Ordnungshomomorphismus.
2. $B \in \mathcal{M}_1$: Damit ist $\Upsilon(B) = B \setminus I \cup \text{Ro}(\mathcal{N})$.
 - (a) $A \in \mathcal{M}'_3$: Damit gilt $\Upsilon(A) = A$ und $I \cap A = \emptyset$. Somit ist $\Upsilon(A) = A \subseteq B \setminus I \subseteq B \setminus I \cup \text{Ro}(\mathcal{N}) = \Upsilon(B)$.
 - (b) $A \in \mathcal{M}_1$: Damit gilt $\Upsilon(A) = A \setminus I \cup \text{Ro}(\mathcal{N}) \subseteq B \setminus I \cup \text{Ro}(\mathcal{N}) = \Upsilon(B)$.

Somit ist Υ ein Ordnungshomomorphismus. Da nach Lemma 4.1 Υ bijektiv ist, folgt Υ ist ein Ordnungsisomorphismus. □

Verw. von: Lem. 4.1; *Verw. in:* Kor. 4.1, 4.2, Lem. 4.5, 9.3.

Aus der Ordnungsisomorphie von Υ folgen die beiden folgenden Korollare 4.1.

Korollar 4.1

Da Υ ein Ordnungsisomorphismus ist, übertragen sich auch \sqcup und \sqcap isomorph zwischen $\mathcal{M}_1 \cup \mathcal{M}'_3$ und $\mathcal{L}_1 \cup \mathcal{L}'_3$.

Verw. von: Lem. 4.2; *Verw. in:* Lem. 9.2, 9.3.

Korollar 4.2

Sei $\mathcal{M}' := \mathcal{M}_1 \cup \mathcal{M}_3$ und $\mathcal{L}' := \mathcal{L}_1 \cup \mathcal{L}_3$. Da Υ ein Ordnungsisomorphismus ist, gilt $\Upsilon(\text{Pa}_{\mathcal{M}'}(A)) = \text{Pa}_{\mathcal{L}'}(\Upsilon(A))$, für jedes $A \in \mathcal{M}'$ und $\Upsilon^{-1}(\text{Pa}_{\mathcal{L}'}(B)) = \text{Pa}_{\mathcal{M}'}(\Upsilon^{-1}(B))$, für jedes $B \in \mathcal{L}'$.

Verw. von: Lem. 4.2; Verw. in: Lem. 9.2, 9.3.

Aus $\mathcal{M}_1 \cup \mathcal{M}_3$ übertragen sich weitere Eigenschaften auf $\mathcal{L}_1 \cup \mathcal{L}_3$.

Lemma 4.3

Sei $X \in \mathcal{M}_1 \cup \mathcal{M}_3$, dann gilt $B(\Upsilon(X)) = B(X)$. Sei $Y \in \mathcal{L}_1 \cup \mathcal{L}_3$, dann gilt $B(\Upsilon^{-1}(Y)) = B(Y)$.

Beweis. Wir zeigen $B(\Upsilon(X)) = B(X)$. Für $X \in \mathcal{M}_3$ ist die Aussage offensichtlich. Sei $A \in \mathcal{M}_1$, dann gilt:

$$\begin{aligned} B(\Upsilon(X)) &= B(X \setminus I \cup \text{Ro}(\mathcal{N})) \\ &= B(X \setminus I) \cup B(\text{Ro}(\mathcal{N})) \\ &= B(X \setminus B(I)) \cup B(I) \quad (\text{da nach Vor. } B(\text{Ro}(\mathcal{B})) = B(I) \text{ gilt}) \\ &= B(X) \quad (\text{da } I \subset X \text{ gilt}) \end{aligned}$$

Wir zeigen analog $B(\Upsilon^{-1}(B)) = B(B)$. Für $Y \in \mathcal{L}_3$ ist die Aussage offensichtlich. Sei $Y \in \mathcal{L}_1$, dann gilt:

$$\begin{aligned} B(\Upsilon(Y)) &= B(Y \setminus \text{Ro}(\mathcal{N}) \cup I) \\ &= B(Y \setminus B(\text{Ro}(\mathcal{N}))) \cup B(I) \\ &= B(X \setminus B(I)) \cup B(\text{Ro}(\mathcal{N})) \quad (\text{da nach Vor. } B(\text{Ro}(\mathcal{B})) = B(I) \text{ gilt}) \\ &= B(X) \quad (\text{da } \text{Ro}(\mathcal{N}) \subset Y \text{ gilt}) \end{aligned}$$

□

Verw. von: Lem. 4.1; Verw. in: Theo. 4.1, Lem. 4.5.

Lemma 4.4

Seien $A, B \in \mathcal{M}_1 \cup \mathcal{M}'_3$, dann gilt $\Upsilon(A \cap B) = \Upsilon(A) \cap \Upsilon(B)$.

Beweis. Da \mathcal{M} ein n -Clustersistem ist, gilt $A \cap B \in \{A, B, \emptyset\}$ für $A, B \in \mathcal{M}_1 \cup \mathcal{M}_3$, woraus dann $A \cap B \in \{A, B, \emptyset\}$ für $A, B \in \mathcal{M}_1 \cup \mathcal{M}'_3$ folgt. Wir betrachten die zwei folgenden Fälle.

1. $A \cap B \in \{A, B\}$: o.B.d.A. gelte $A \cap B = A$, woraus $B \subseteq A$ folgt. Da Υ ein Ordnungsisomorphismus (Lem. 4.2) ist, folgt $\Upsilon(B) \subseteq \Upsilon(A)$, womit $\Upsilon(B) \cap \Upsilon(A) = \Upsilon(A) = \Upsilon(A \cap B)$ gilt.

2. $A \cap B = \emptyset$: Da Υ ein Ordnungsisomorphismus (Lem. 4.2) ist, sind auch $\Upsilon(A)$ und $\Upsilon(B)$ nicht vergleichbar. Also gilt $\Upsilon(A) \cap \Upsilon(B) = \emptyset = \Upsilon(\emptyset)$.

□

Verw. von: Lem. 4.2; Verw. in: Theo. 4.1.

Theorem 4.1 (\mathcal{L} ist ein m -Clustersystem)

\mathcal{L} ist ein m -Clustersystem über $\text{Ro}(\mathcal{M}) \setminus I \cup \text{Ro}(\mathcal{N})$.

Beweis. Wir betrachten zuerst den Fall $I = \text{Ro}(\mathcal{M})$. Damit wird \mathcal{M} vollständig durch \mathcal{N} ersetzt, und es ist $\mathcal{L} = \mathcal{N}$. Somit ist \mathcal{L} ein m -Clustersystem. Wir zeigen nun für $I \neq \text{Ro}(\mathcal{M})$, daß \mathcal{L} ein m -Clustersystem ist. Dazu zeigen wir die folgenden Aussagen.

1. $\emptyset \notin \mathcal{L}$: Es gilt $\mathcal{L} = \Upsilon(\mathcal{M}_1 \cup \mathcal{M}_2) \cup \mathcal{N}$. Für das m -Clustersystem \mathcal{N} gilt $\emptyset \notin \mathcal{N}$. Da nach Lemma 4.1 Υ bijektiv ist, $\Upsilon(\emptyset) = \emptyset$ und $\emptyset \notin \mathcal{M}$, folgt $\emptyset \notin \Upsilon(\mathcal{M}_1 \cup \mathcal{M}_2)$. Somit ist $\emptyset \notin \mathcal{L}$.
2. $\text{Ro}(\mathcal{M}) \setminus I \cup \text{Ro}(\mathcal{N}) \in \mathcal{L}$: Nach Voraussetzung gilt $I \neq \text{Ro}(\mathcal{M})$, woraus $\text{Ro}(\mathcal{M}) \in \mathcal{M}_1$ folgt. Also ist $\Upsilon(\text{Ro}(\mathcal{M})) \in \mathcal{L}$. Somit gilt $\Upsilon(\text{Ro}(\mathcal{M})) = \text{Ro}(\mathcal{M}) \setminus I \cup \text{Ro}(\mathcal{N}) \in \mathcal{L}$.
3. Für jedes $i \in \text{Ro}(\mathcal{M}) \setminus I \cup \text{Ro}(\mathcal{N})$ ist $\{i\} \in \mathcal{L}$: Wir unterscheiden die zwei folgenden Fälle:
 - (a) $i \in \text{Ro}(\mathcal{N})$: \mathcal{N} ist nach Voraussetzung ein m -Clustersystem, und es folgt $\{i\} \in \mathcal{N}$. Somit ist $\{i\} \in \mathcal{L}$.
 - (b) $i \in \text{Ro}(\mathcal{M}) \setminus I$: Wir zeigen $\{i\} \in \mathcal{M}_3$, da hiermit $\{i\} = \Upsilon(\{i\})$ gilt, woraus $\{i\} \in \mathcal{L}$ folgt. \mathcal{M} ist ein m -Clustersystem, womit $\{i\} \in \mathcal{M}$ gilt. Aus $i \notin I$ folgt $\{i\} \notin T_{\mathcal{M}}(I) = \mathcal{M}_2$. Es gilt $\{i\} \in \text{Le}(\mathcal{M})$, woraus $\{i\} \notin K_{\mathcal{M}}(\text{Ro}(\mathcal{M}), \text{Pa}(I)) = \mathcal{M}_1$ folgt. Somit ist $\{i\} \in \mathcal{M}_3$.
4. Für alle $A, B \in \mathcal{L}$ gilt $A \cap B \in \{A, B, \emptyset\}$: Wir unterscheiden die folgenden Fälle:
 - (a) $A, B \in \mathcal{L}_1 \cap \mathcal{L}_3$: Da \mathcal{M} ein m -Clustersystem ist und A, B unter Υ^{-1} nach \mathcal{M} abbilden, gilt $\Upsilon^{-1}(A) \cap \Upsilon^{-1}(B) \in \{\Upsilon^{-1}(A), \Upsilon^{-1}(B), \emptyset\}$. Damit folgt, $\Upsilon(\Upsilon^{-1}(A) \cap \Upsilon^{-1}(B)) \in \Upsilon(\{\Upsilon^{-1}(A), \Upsilon^{-1}(B), \emptyset\})$. Mit Lemma 4.4 folgt $\Upsilon(\Upsilon^{-1}(A) \cap \Upsilon^{-1}(B)) = \Upsilon(\Upsilon^{-1}(A \cap B)) = A \cap B$. Somit folgt $A \cap B \in \{A, B, \emptyset\}$.

(b) $A \in \mathcal{L}_2$:

- i. $B \in \mathcal{L}_2$: Da $\mathcal{N} = \mathcal{L}_2$ und \mathcal{N} ein m -Clustersystem ist, gilt die Aussage.
- ii. $B \in \mathcal{L}_3$: Nach Voraussetzung sind \mathcal{M} und \mathcal{N} disjunkt. Somit gilt $A \cap B = \emptyset$.
- iii. $B \in \mathcal{L}_1$: Also gilt $A \subset B$ und somit $A \cap B = B$.

5. Sei $C \in \mathcal{L}$ und $A, B \in \text{Ch}(C)$ mit $A \neq B$ und $B(A) \cap B(B) \neq \emptyset$, dann gilt $B(A) = B(B) = B(C)$. Wir zeigen dies durch die folgenden Fälle:

(a) $A, B \in \mathcal{L}_1 \cup \mathcal{L}_3$:

Erstens gilt $A \neq B$, woraus $\Upsilon^{-1}(A) \neq \Upsilon^{-1}(B)$ folgt, da Υ ein Isomorphismus ist. Zweitens gilt $B(A) \cap B(B) \neq \emptyset$, also $B(\Upsilon(\Upsilon^{-1}(A))) \cap B(\Upsilon(\Upsilon^{-1}(B))) \neq \emptyset$, woraus $B(\Upsilon^{-1}(A)) \cap B(\Upsilon^{-1}(B)) \neq \emptyset$ mit Lemma 4.3 folgt. Drittens gilt $A, B \in \text{Ch}(C)$, woraus $\Upsilon^{-1}(A), \Upsilon^{-1}(B) \in \text{Ch}(\Upsilon^{-1}(C))$ nach Korollar 4.2 folgt. Da \mathcal{M} ein m -Clustersystem ist, folgt aus diesen drei Aussagen $B(\Upsilon^{-1}(A)) = B(\Upsilon^{-1}(B)) = B(\Upsilon^{-1}(C))$.

Mit Lemma 4.3 folgt somit $B(A) = B(B) = B(C)$.

(b) $A \in \mathcal{L}_2$:

- i. $B \in \mathcal{L}_2$: Gilt offensichtlich, da $\mathcal{L}_2 (= \mathcal{N})$ nach Voraussetzung ein m -Clustersystem ist.
- ii. $B \in \mathcal{L}_3$: Erstens gilt $A \neq B$, woraus $\Upsilon^{-1}(A) \neq \Upsilon^{-1}(B)$ folgt, da Υ ein Isomorphismus ist. Zweitens folgt $A = \text{Ro}(\mathcal{N})$ und $C = \text{Pa}(\text{Ro}(\mathcal{N}))$ aus der Voraussetzung $A, B \in \text{Ch}(C)$. Für $A = \text{Ro}(\mathcal{N})$ gilt $B(A) = B(I)$, nach der Voraussetzung für die Ersetzung \mathcal{L} . Also gilt $B(I) \cap B(B) = B(A) \cap B(B) \neq \emptyset$. Nehmen wir $I, B \in \text{Ch}_{\mathcal{M}}(\Upsilon^{-1}(C))$ an. Da \mathcal{M} ein m -Clustersystem ist, folgt damit aus den beiden vorangegangenen Aussagen $B(I) = B(B) = B(\Upsilon^{-1}(C))$. Es gilt $B(I) = B(A)$ und nach Lemma 4.3 gilt $B(\Upsilon^{-1}(C)) = B(C)$. Damit folgt $B(A) = B(B) = B(C)$. Wir zeigen $B \in \text{Ch}_{\mathcal{M}}(\Upsilon^{-1}(C))$. Aus Korollar 4.2 folgt $\Upsilon^{-1}(\text{Pa}_{\mathcal{L}}(B)) = \text{Pa}_{\mathcal{M}}(\Upsilon^{-1}(B))$. Es gilt $\Upsilon^{-1}(B) = B$, da $B \in \mathcal{L}_3$. Somit folgt $\Upsilon^{-1}(C) = \Upsilon^{-1}(\text{Pa}_{\mathcal{L}}(B)) = \text{Pa}_{\mathcal{M}}(B)$. Wir zeigen $I \in \text{Ch}_{\mathcal{M}}(\Upsilon^{-1}(C))$. Da Υ ein Ordnungsisomorphismus ist, gilt $\Upsilon^{-1}(\min(\mathcal{L}_1)) = \min(\mathcal{M}_1)$. Aus $C = \text{Pa}(\text{Ro}(\mathcal{N}))$ folgt $\Upsilon^{-1}(\min(\mathcal{L}_1)) = C$ und nach Definition von \mathcal{M}_1 gilt $\min(\mathcal{M}_1) = \text{Pa}_{\mathcal{M}}(I)$. Somit ist $\text{Pa}_{\mathcal{M}}(I) = C$.
- iii. $B \in \mathcal{L}_1$: Damit würde $A \subset B$ folgen und der Voraussetzung $A, B \in \text{Ch}(C)$ widersprechen.

□

4.1.2 Die Operation: Merge

Die Operation $\text{Merge}(\mathcal{N}_1, \dots, \mathcal{N}_n)$ fügt, wie in Abbildung 4.3 gezeigt ist, m -Clustersysteme mit Wurzeln gleicher Basis zu einem neuen m -Clustersystem zusammen.

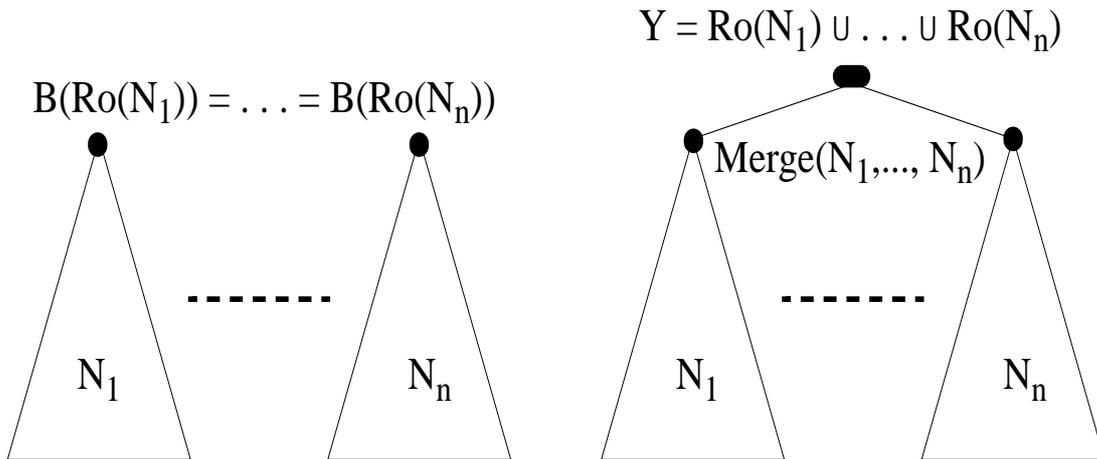


Abbildung 4.3: Die Zusammenfügung $\text{Merge}(\mathcal{N}_1, \dots, \mathcal{N}_n)$

$\text{Merge}(\mathcal{N}_1, \dots, \mathcal{N}_n)$ entsteht durch Vereinigung von \mathcal{N}_1 bis \mathcal{N}_n und dem neuen gemeinsamen Elternelement Y aller Wurzeln der m -Clustersysteme. Damit bei der Vereinigung von \mathcal{N}_1 bis \mathcal{N}_n die Baumeigenschaft von $\text{Merge}(\mathcal{N}_1, \dots, \mathcal{N}_n)$ nicht "verhindert" wird, fordern wir die paarweise Disjunktheit der Wurzeln dieser m -Clustersysteme.

Definition 4.6 (Merge)

Sei $\{\mathcal{N}_1, \dots, \mathcal{N}_n\}$ eine Menge von m -Clustersysteme mit $n \in \mathbb{N}$, für die folgendes gilt: für alle $\mathcal{X}, \mathcal{Y} \in \{\mathcal{N}_1, \dots, \mathcal{N}_n\}$ mit $\mathcal{X} \neq \mathcal{Y}$ gilt $B(\text{Ro}(\mathcal{X})) = B(\text{Ro}(\mathcal{Y}))$ und $\text{Ro}(\mathcal{N}_i) \cap \text{Ro}(\mathcal{N}_j) = \emptyset$.

Wir definieren $Y := \bigcup_{i \in \{1, \dots, n\}} \text{Ro}(\mathcal{N}_i)$ und bezeichnen

$$\text{Merge}(\mathcal{N}_1, \dots, \mathcal{N}_n) := Y \cup \bigcup_{i \in \{1, \dots, n\}} \mathcal{N}_i$$

als die Zusammenfügung der m -Clustersysteme \mathcal{N}_1 bis \mathcal{N}_n . Sind die Komponenten der Zusammenfügung offensichtlich, so verwenden wir die Bezeichnung Zusammenfügung.

Theorem 4.2 (Merge ist ein m -Clustersystem)

Die Zusammenfügung $\text{Merge}(\mathcal{N}_1, \dots, \mathcal{N}_n)$ ist ein m -Clustersystem über der Multimenge Y .

Beweis. Wir zeigen die folgenden Eigenschaften des m -Clustersystemes. Zur Vereinfachung sei $\mathcal{M}e := \text{Merge}(\mathcal{N}_1, \dots, \mathcal{N}_n)$.

1. $\emptyset \notin \mathcal{M}e$: gilt offensichtlich.
2. $Y \in \mathcal{M}e$: gilt nach Definition von $\mathcal{M}e$.
3. Für jedes $i \in Y$ gilt $\{i\} \in \mathcal{M}e$: Da $Y = \bigcup_{k \in \{1, \dots, n\}} \text{Ro}(\mathcal{N}_k)$, existiert ein \mathcal{N}_j mit $i \in \text{Ro}(\mathcal{N}_j)$ und $j \in \{1, \dots, n\}$. Da \mathcal{N}_j ein m -Clustersystem ist, gilt $\{i\} \in \mathcal{N}_j$.
4. Für alle $A, B \in \mathcal{M}e$ gilt $A \cap B \in \{A, B, \emptyset\}$: Wir unterscheiden o.B.d.A. die folgenden Fälle.
 - (a) es gibt ein $i \in \{1, \dots, n\}$ mit
 - i. $A, B \in \mathcal{N}_i$: $A \cap B \in \{A, B, \emptyset\}$ gilt, da \mathcal{N}_i ein m -Clustersystem ist.
 - ii. $A \in \mathcal{N}_i$ und $A = Y$: Damit gilt $B \subset A$, woraus $A \cap B = A$ folgt.
 - (b) es gibt ungleiche $i, j \in \{1, \dots, n\}$ mit $A \in \mathcal{N}_i$ und $B \in \mathcal{N}_j$: Da $\mathcal{M}e$ eine Zusammenfügung ist, gilt $\text{Ro}(\mathcal{N}_i) \cap \text{Ro}(\mathcal{N}_j) = \emptyset$, womit $A \cap B = \emptyset$ folgt.
 - (c) $A = B = Y$: klar.
5. Sei $C \in \mathcal{M}e$ und $A, B \in \text{Ch}(C)$ mit $A \neq B$ und $B(A) \cap B(B) \neq \emptyset$, dann gilt $B(A) = B(B) = B(C)$: Wir unterscheiden die beiden folgenden Fälle.
 - (a) es gibt ein $i \in \{1, \dots, n\}$ mit $C \in \mathcal{N}_i$: es folgt $A, B \in \mathcal{N}_i$. Da \mathcal{N}_i ein m -Clustersystem ist, gilt die Behauptung.
 - (b) $C = Y$: Damit existieren ungleiche $i, j \in \{1, \dots, n\}$ mit $A = \text{Ro}(\mathcal{N}_i)$ und $B = \text{Ro}(\mathcal{N}_j)$. Es gilt $Y = \bigcup_{i \in \{1, \dots, n\}} \text{Ro}(\mathcal{N}_i)$. Da die Wurzeln aller m -Clustersysteme $\mathcal{N}_1, \dots, \mathcal{N}_n$ die gleiche Basis besitzen, folgt die Behauptung.

□

Verw. von: ; Verw. in: Theo. 4.4.

Ist das m -Clustersystem $\text{Merge}(\mathcal{N}_1, \dots, \mathcal{N}_n)$ von einem Speziesbaum ableitbar, so ist dessen Wurzel Y ein Duplikationselement. Damit ist Y ein Duplikationselement, wenn die m -Clustersysteme $\mathcal{N}_1, \dots, \mathcal{N}_n$ von demselben Speziesbaum ableitbar sind.

4.1.3 Die Operation: \times

Durch die Operation \times erzeugen wir aus einem n -Clustersistem ein m -Clustersistem, indem wir die Mengen des n -Clustersistemes als Multimengen interpretieren.

Definition 4.7 (\times)

Sei \mathcal{S} ein n -Clustersistem und $p \in \mathbb{N}$, dann ist $\mathcal{S} \times p := \{\mathcal{S} \times \{p\} \mid \exists \mathcal{S} \in \mathcal{S}\}$.

Theorem 4.3 ($\mathcal{S} \times p$ ist ein m -Clustersistem)

Sei \mathcal{S} ein n -Clustersistem und $p \in \mathbb{N}$. Dann ist $\mathcal{S} \times p$ ein m -Clustersistem über der Multimenge $\text{Ro}(\mathcal{S}) \times \{p\}$.

Beweis. $\mathcal{S} \times p$ ist ein n -Clustersistem, da \mathcal{S} ein n -Clustersistem ist und sich alle Eigenschaften des n -Clustersistemes auf $\mathcal{S} \times p$ übertragen.

Wir zeigen: Sei $C \in \mathcal{S} \times p$, $A, B \in \text{Ch}_{\mathcal{S}}(C)$ mit $A \neq B$ und $B(A) \cap B(B) \neq \emptyset$, dann gilt $B(A) = B(B) = B(C)$.

Da $A, B, C \in \mathcal{S} \times p$ sind, existieren $A', B', C' \in \mathcal{S}$ mit $A = A' \times \{p\}$, $B = B' \times \{p\}$ und $C = C' \times \{p\}$. Aus $A, B \in \text{Ch}_{\mathcal{S}}(C)$ mit $A \neq B$ folgt $A' \neq B'$ sowie $A', B' \in \text{Ch}_{\mathcal{S}}(C')$. Es gilt $A' \cap B' \in \{A, B, \emptyset\}$, da \mathcal{S} ein n -Clustersistem ist. Wegen $A' \neq B'$ und $A', B' \in \text{Ch}_{\mathcal{S}}(C')$ folgt $A' \cap B' = \emptyset$.

Somit gilt $B(A) \cap B(B) = A' \cap B' = \emptyset$, womit $\mathcal{S} \times p$ keine Duplikationselemente beinhaltet. \square

Verw. von: ; Verw. in: Lem. 9.1.

Die Umbenennung durch \times erzeugt einen ohne Duplikationen aus \mathcal{S} abgeleiteten Genbaum $\mathcal{S} \times p$, mit $p \in \mathbb{N}_0$. Wollen wir einen Teilbaum in einem m -Clustersistem \mathcal{M} durch einen ohne Duplikationen abgeleiteten Teilbaum $\mathcal{S}' \times p$ ersetzen, so muß $\text{Ro}(\mathcal{S}' \times p) \cap \text{Ro}(\mathcal{M}) = \emptyset$ gelten. Dies erreichen wir, indem wir p (tdisjunkt zu $\text{Ro}(\mathcal{M})$) so wählen, daß es in keinem Zweitupel von $\text{Ro}(\mathcal{M})$ vorkommt.

Definition 4.8 (tdisjunkt)

Wir bezeichnen $p \in \mathbb{N}$ tdisjunkt zu einer Menge M , wenn p in keinem Zweitupel von M vorkommt.

4.2 Die Ableitung von Genbäumen

Durch die Operationen *Repl* und *Merge* können wir jeweils, unter gewissen Voraussetzungen, aus spezieskonformen Ableitungen eines Speziesbaumes \mathcal{S} weitere

Ableitungen von \mathcal{S} erzeugen. Diese Eigenschaft verwenden wir, um in Abschnitt 4.3 und 4.4 Reconciled Trees aus einem Speziesbaum abzuleiten.

Ersetzen wir einen von \mathcal{S} abgeleiteten Genbaum \mathcal{N} in einem von \mathcal{S} abgeleiteten Genbaum \mathcal{M} an einem $I \in \mathcal{M}$, so erhalten wir wieder einen von \mathcal{S} abgeleiteten Genbaum.

Lemma 4.5

Sei $\mathcal{S} \in \mathfrak{N}(\mathcal{M})$ über einer Menge M , $\mathcal{M}' := \text{Repl}(\mathcal{M}, \mathcal{N}, I)$ eine Ersetzung, wobei \mathcal{M} und \mathcal{N} von \mathcal{S} ableitbar sind. Dann ist \mathcal{M}' von \mathcal{S} ableitbar.

Beweis.

Ist $I = \text{Ro}(\mathcal{M})$, so gilt die Behauptung, da $\mathcal{M}' = \mathcal{N}$ folgt und \mathcal{N} nach Voraussetzung von \mathcal{S} ableitbar ist. Betrachten wir nun den Fall $I \neq \text{Ro}(\mathcal{M})$.

\mathcal{M}' ist nach Voraussetzung eine Ersetzung, womit die Funktion $\Upsilon := \Upsilon_{\mathcal{M}, \mathcal{N}, I}$ existiert. Seien $X, Y \in \mathcal{M}'$ mit $Y \subseteq X$. Wir zeigen o.B.d.A. die Behauptung $B(K'(X, Y)) = K_S(B(X), B(Y))$ anhand der folgenden Fälle:

1. $X, Y \in \overline{T_{\mathcal{M}'}}(\text{Ro}(\mathcal{N}))$:

$$B(K_{\mathcal{M}'}(X, Y)) = B(\Upsilon^{-1}(K_{\mathcal{M}'}(X, Y))) \quad (4.1)$$

$$= B(K_{\mathcal{M}}(\Upsilon^{-1}(X), \Upsilon^{-1}(Y))) \quad (4.2)$$

$$= K_S(B(\Upsilon^{-1}(X)), B(\Upsilon^{-1}(Y))) \quad (4.3)$$

$$= K_S(B(X), B(Y)). \quad (4.4)$$

Zu (4.1), (4.4): wie in Lemma 4.3 gezeigt, gilt $B(\Upsilon^{-1}(X)) = B(X)$ für alle $X \in \overline{T_{\mathcal{M}'}}(\text{Ro}(\mathcal{N}))$.

Zu (4.2): nach Lemma 4.2 ist Υ ein Ordnungsisomorphismus.

Zu (4.3): \mathcal{M} ist nach Voraussetzung von \mathcal{S} ableitbar.

2. $X, Y \in T_{\mathcal{M}'}(\mathcal{N})$; nach Voraussetzung ist \mathcal{N} von \mathcal{S} ableitbar.

3. $X \in \overline{T_{\mathcal{M}'}}(\text{Ro}(\mathcal{N}))$, $Y \in T_{\mathcal{M}'}(\text{Ro}(\mathcal{N}))$:

$$\begin{aligned} B(K_{\mathcal{M}'}(X, Y)) &= B(K_{\mathcal{M}'}(X, \text{Pa}_{\mathcal{M}'}(\text{Ro}(\mathcal{N}))) \cup K_{\mathcal{M}'}(\text{Ro}(\mathcal{N}), Y)) \\ &= B(K_{\mathcal{M}'}(X, \text{Pa}_{\mathcal{M}'}(\text{Ro}(\mathcal{N})))) \cup B(K_{\mathcal{M}'}(\text{Ro}(\mathcal{N}), Y)) \\ &= K_S(B(X), B(\text{Pa}_{\mathcal{M}'}(\text{Ro}(\mathcal{N})))) \\ &\quad \cup K_S(B(\text{Ro}(\mathcal{N})), B(Y)) \end{aligned} \quad (4.5)$$

$$\begin{aligned} &= K_S(B(X), B(\text{Pa}_{\mathcal{M}'}(\text{Ro}(\mathcal{N})))) \\ &\quad \cup K_S(B(\text{Pa}_{\mathcal{M}'}(\text{Ro}(\mathcal{N}))), B(\text{Ro}(\mathcal{N}))) \\ &\quad \cup K_S(B(\text{Ro}(\mathcal{N})), B(Y)) \end{aligned} \quad (4.6)$$

$$= K_S(B(X), B(Y))$$

Zu (4.5): folgt direkt aus den Aussagen 1 und 2.

Zu (4.6): Wir zeigen

$$K_S(\mathcal{B}(\text{Pa}_{\mathcal{M}'}(\text{Ro}(\mathcal{N}))), \mathcal{B}(\text{Ro}(\mathcal{N}))) = \{\mathcal{B}(\text{Pa}_{\mathcal{M}'}(\text{Ro}(\mathcal{N}))), \mathcal{B}(\text{Ro}(\mathcal{N}))\}.$$

Da \mathcal{M}' eine Ersetzung von \mathcal{N} an Element I in \mathcal{M} ist, gilt $\Upsilon^{-1}(\text{Pa}_{\mathcal{M}'}(\text{Ro}(\mathcal{N}))) = \text{Pa}_{\mathcal{M}}(I)$, und es folgt:

$$\mathcal{B}(\text{Pa}_{\mathcal{M}'}(\text{Ro}(\mathcal{N}))) = \mathcal{B}(\Upsilon^{-1}(\text{Pa}_{\mathcal{M}'}(\text{Ro}(\mathcal{N})))) = \mathcal{B}(\text{Pa}_{\mathcal{M}}(I)).$$

Da \mathcal{M}' eine Ersetzung ist, gilt $\mathcal{B}(\text{Ro}(\mathcal{N})) = \mathcal{B}(I)$ und nach Voraussetzung für \mathcal{M} gilt $K_S(\mathcal{B}(\text{Pa}_{\mathcal{M}}(I)), \mathcal{B}(I)) = \mathcal{B}(K_{\mathcal{M}}(\text{Pa}_{\mathcal{M}}(I), I))$. Hieraus folgt:

$$\begin{aligned} K_S(\mathcal{B}(\text{Pa}_{\mathcal{M}'}(\text{Ro}(\mathcal{N}))), \mathcal{B}(\text{Ro}(\mathcal{N}))) &= K_S(\mathcal{B}(\text{Pa}_{\mathcal{M}}(I)), \mathcal{B}(I)) \\ &= \mathcal{B}(K_{\mathcal{M}}(\text{Pa}_{\mathcal{M}}(I), I)) \\ &= \{\mathcal{B}(\text{Pa}_{\mathcal{M}}(I)), \mathcal{B}(I)\} \\ &= \{\mathcal{B}(\text{Pa}_{\mathcal{M}'}(\text{Ro}(\mathcal{N}))), \mathcal{B}(\text{Ro}(\mathcal{N}))\}. \end{aligned}$$

□

Verw. von: Lem. 4.2, 4.3; Verw. in: Lem. 9.2, 2, Theo 4.4.

Fügen wir die von einem Teilbaum von \mathcal{S} abgeleiteten Genbäume $\mathcal{N}_1, \dots, \mathcal{N}_n$ zu $\mathcal{M}' = \text{Merge}(\mathcal{N}_1, \dots, \mathcal{N}_n)$ zusammen, so ist \mathcal{M}' wieder ein von \mathcal{S} abgeleiteter Genbaum.

Lemma 4.6

Sei $\mathcal{S} \in \mathfrak{N}(M)$ über einer Menge M . $\mathcal{M}' := \text{Merge}(\mathcal{N}_1, \dots, \mathcal{N}_n)$ sei eine Zusammenfügung mit $n \in \mathbb{N}$, wobei \mathcal{N}_i für jedes $i \in \{1, \dots, n\}$ von \mathcal{S} ableitbar ist. Dann ist \mathcal{M}' von \mathcal{S} ableitbar.

Beweis.

Seien $X, Y \in \mathcal{M}'$ mit $Y \subseteq X$. Wir zeigen die Behauptung für die folgenden Fälle:

1. $X, Y \in T_{\mathcal{M}'}(\text{Ro}(\mathcal{N}_i))$ für ein $i \in \{1, \dots, n\}$: gilt nach Voraussetzung.
2. $X = Y = \text{Ro}(\mathcal{M}')$: nach Definition der Zusammenfügung gilt $\text{Ro}(\mathcal{M}') = \bigcup_{i \in \{1, \dots, n\}} \text{Ro}(\mathcal{N}_i)$ und $\mathcal{B}(\mathcal{N}_i) = \mathcal{B}(\mathcal{N}_j)$ für alle $i, j \in \{1, \dots, n\}$. Für ein beliebiges $i \in \{1, \dots, n\}$ folgt damit $\mathcal{B}(\text{Ro}(\mathcal{M}')) = \mathcal{B}(\text{Ro}(\mathcal{N}_i))$. Da nach Voraussetzung $\mathcal{B}(\text{Ro}(\mathcal{N}_i)) \in \mathcal{S}$ gilt, folgt $\mathcal{B}(X) \in \mathcal{S}$. Somit gilt $\mathcal{B}(K_{\mathcal{M}'}(X, Y)) = \mathcal{B}(X) = K_S(\mathcal{B}(X), \mathcal{B}(Y))$.

3. $X = \text{Ro}(\mathcal{M}')$, $Y \in \text{T}_{\mathcal{M}'}(\text{Ro}(\mathcal{N}_i))$ für ein $i \in \{1, \dots, n\}$: nach Definition der Zusammenfügung gilt $X = \text{Pa}_{\mathcal{M}'}(\text{Ro}(\mathcal{N}_i))$. Der Fall 2 zeigt, daß $B(X) = B(\text{Ro}(\mathcal{N}_i))$ gilt. Aus diesen beiden Aussagen folgt:

$$B(K_{\mathcal{M}'}(X, Y)) = B(\{X\} \cup K_{\mathcal{M}'}(\text{Ro}(\mathcal{N}_i), Y)) = B(K_{\mathcal{M}'}(\text{Ro}(\mathcal{N}_i), Y)).$$

Nach Voraussetzung gilt $B(K_{\mathcal{M}'}(\text{Ro}(\mathcal{N}_i), Y)) = K_S(B(\text{Ro}(\mathcal{N}_i)), B(Y))$, und es folgt mit $B(X) = B(\text{Ro}(\mathcal{N}_i))$ die Behauptung.

4. $X \in \text{T}_{\mathcal{M}'}(\text{Ro}(\mathcal{N}_i))$, $Y \in \text{T}_{\mathcal{M}'}(\text{Ro}(\mathcal{N}_j))$ für ungleiche $i, j \in \{1, \dots, n\}$: damit sind X, Y nicht vergleichbar, womit dieser Fall nicht vorkommt.

□

Verw. von: ; Verw. in: Theo. 4.4.

4.3 Das LCA-Theorem

Betrachten wir einen Reconciled Tree $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$ des Genbaumes \mathcal{G} über dem Speziesbaum \mathcal{S} . Nach der Definition des Reconciled Tree existiert ein spezieskonformer Einbettungsmonomorphismus der Menge $\text{Emon}(\mathcal{G}, \mathcal{R})$ des Genbaumes \mathcal{G} in den Speziesbaum \mathcal{S} . Aus der Menge $\text{Emon}(\mathcal{G}, \mathcal{R})$ betrachten wir nun einen spezieskonformen Einbettungsmonomorphismus, welcher die Elemente aus \mathcal{G} auf die von heute aus frühest möglichen Gene in \mathcal{R} abbildet. Einen solchen Einbettungsmonomorphismus bezeichnen wir nach Definition 9.9 als minimal.

Sei $\varphi \in \text{Emon}(\mathcal{G}, \mathcal{R})$ ein minimaler spezieskonformer Einbettungsmonomorphismus. Wir wollen nun eine Aussage über die Spezies der Gene in \mathcal{R} machen, in welche der Genbaum eingebettet ist, also über die Menge $\varphi(\mathcal{G})$. Die Spezies eines Genes $\varphi(X) \in \mathcal{R}$ mit $X \in \mathcal{G}$ ist durch die Basis $B(\varphi(X))$, und die Host-Spezies von X ist durch die kleinste oberer Schranke $\sqcup_{\mathcal{S}}(X)$ gegeben. Nach dem LCA-Theorem ist nun die Host-Spezies von X gleich mit der Spezies von X . Allgemein besagt das LCA-Theorem, daß $B(\varphi(X)) = \sqcup_{\mathcal{S}}(X)$ für alle $X \in \mathcal{G}$ gilt. Für eine zwar spezieskonforme, aber nicht minimale Einbettung muß, wie man sich an einem Beispiel leicht klarmacht, diese Aussage nicht gelten.

Der formale Beweis des LCA-Theoremes ist in Abschnitt 4.3.2 (auf Seite 56) gezeigt. Einen Überblick zu diesem Beweis geben wir vorab in dem nächsten Abschnitt 4.3.1.

4.3.1 Beweisskizze

Für ein $X \in \mathcal{G}$ ist nach Lemma 3.4 (Seite 32) die Host-Spezies $\sqcup_{\mathcal{S}}(X)$ eine untere Schranke $\sqcup_{\mathcal{S}}(X) \subseteq B(\varphi(X))$ für $B(\varphi(X))$. Wir zeigen indirekt, daß $\sqcup_{\mathcal{S}}(X) = B(\varphi(X))$ gilt.

Hierzu nehmen wir die Existenz eines Elementes $W \in \mathcal{G}$ mit $\sqcup_{\mathcal{S}}(W) \neq B(\varphi(W))$ an, womit $\sqcup_{\mathcal{S}}(W) \subset B(\varphi(W))$ gilt. Unter dieser Annahme konstruieren wir aus \mathcal{R} einen Reconciled Tree \mathcal{R}'' mit geringerer Kardinalität als \mathcal{R} , dessen Existenz jedoch im Widerspruch zu der vorausgesetzten minimalen Kardinalität von \mathcal{R} steht.

Der indirekte Beweis unterteilt sich in den strukturellen Abschnitt 4.3.1.1 und konstruktiven Abschnitt 4.3.1.2. In dem strukturellen Abschnitt leiten wir aus \mathcal{R}, \mathcal{S} und \mathcal{G} Aussagen ab, mit deren Hilfe wir in dem konstruktiven Abschnitt aus \mathcal{R} dann den Reconciled Tree \mathcal{R}'' konstruieren.

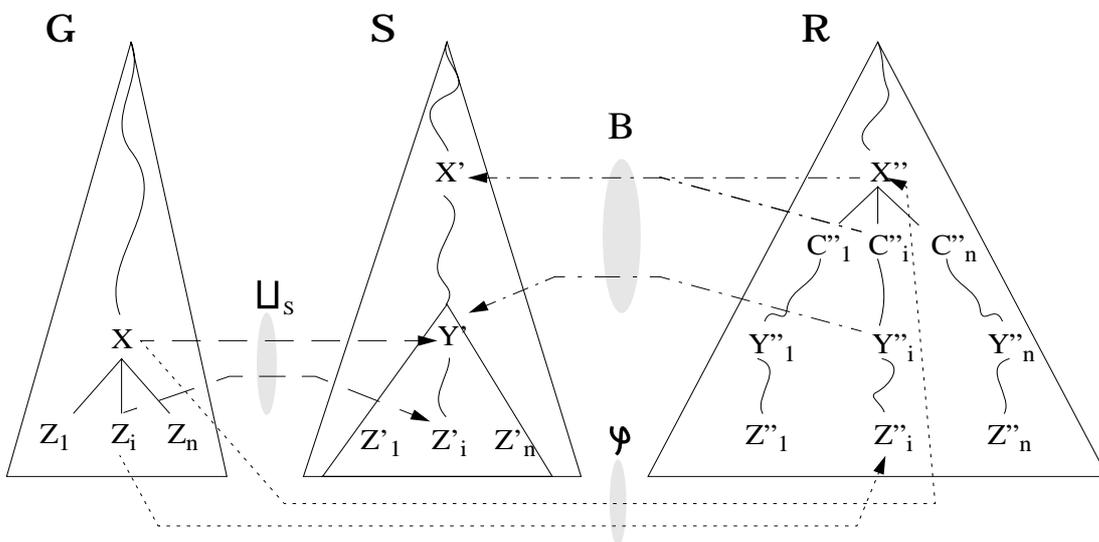


Abbildung 4.4: Strukturelle Eigenschaften

4.3.1.1 Struktureller Abschnitt

In dem strukturellen Abschnitt (siehe Abbildung 4.4) zeigen wir die Existenz eines Elementes $X \in \mathcal{G}$, mit $\sqcup_{\mathcal{S}}(X) \subset B(\varphi(X))$, so daß $\sqcup_{\mathcal{S}}(V) = B(\varphi(V))$, für alle $V \in \mathcal{R}$ mit $V \subset X$ gilt. Diese Eigenschaft von X bezeichnen wir als minimal bezüglich $\sqcup_{\mathcal{S}}$. Aus dieser Eigenschaft folgern wir, daß X Kinder besitzt, welche wir mit Z_1, \dots, Z_n bezeichnen. Die Bilder von X sowie Z_1, \dots, Z_n unter φ bezeichnen wir zur Vereinfachung jeweils mit X' sowie Z'_1, \dots, Z'_n .

X und ein beliebiges Kind Z_i beschreiben in \mathcal{R} die Kette $K_{\mathcal{R}}(X'', Z_i'')$. Die Basis $B(K_{\mathcal{R}}(X'', Z_i''))$ dieser Kette ist nach Voraussetzung $B(K_{\mathcal{R}}(X'', Z_i'')) = K_S(B(X''), B(Z_i''))$ eine Kette in S . X haben wir minimal bezüglich \sqcup_S gewählt, womit $\sqcup_S(Z_i) = B(Z_i'')$ und $\sqcup_S(X) \subset B(X'')$ gilt. Also gilt $\sqcup_S(X) \in B(K_{\mathcal{R}}(X'', Z_i''))$ mit $\sqcup_S(X) \neq B(X'')$. Da $B(K_{\mathcal{R}}(X'', Z_i'')) = K_S(B(X''), B(Z_i''))$ gilt, existiert ein $Y_i'' \in K_{\mathcal{R}}(X'', Z_i'')$ mit $B(Y_i'') = \sqcup_S(X)$ und $Y_i'' \neq X''$. Es kann mehrere Elemente mit dieser Eigenschaft geben. Wir legen Y_i'' als das maximale Element mit dieser Eigenschaft fest.

Betrachten wir nun zwei unterschiedliche Elemente $Y_i'' \in K_{\mathcal{R}}(X'', Z_i'')$ und $Y_j'' \in K_{\mathcal{R}}(X'', Z_j'')$ mit der genannten Eigenschaft. Wir wissen, daß $B(Y_i'') = B(Y_j'') = \sqcup_S(X)$ gilt und damit die Ketten $B(K_{\mathcal{R}}(X'', Y_i''))$ und $B(K_{\mathcal{R}}(X'', Y_j''))$ identisch sind. Mit $Y_i'' \subset X''$ und $Y_j'' \subset X''$ folgt, daß X'' nach Definition 3.7 (Seite 32) ein Duplikationselement ist.

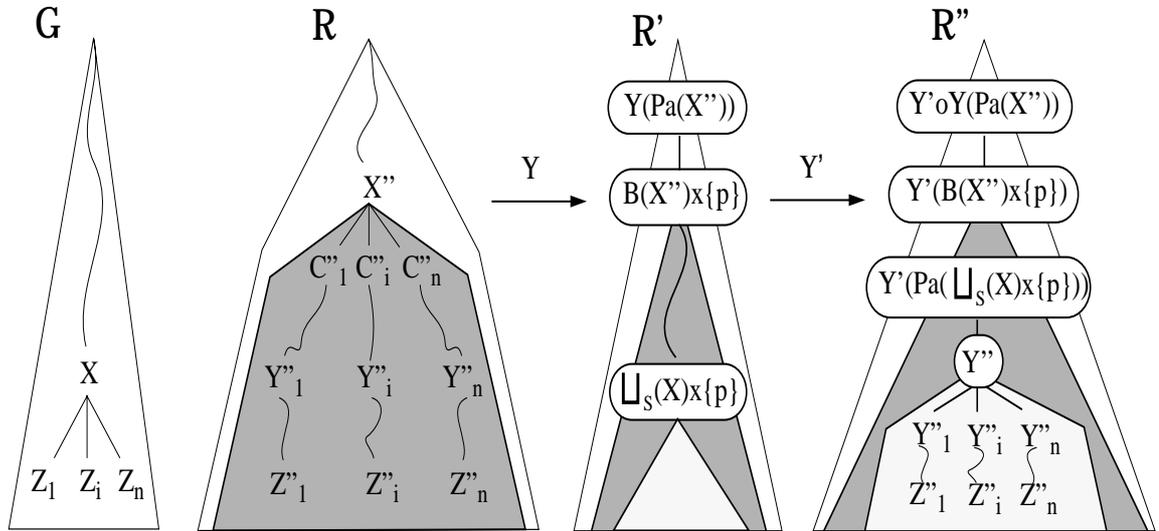


Abbildung 4.5: Konstruktion von \mathcal{R}''

4.3.1.2 Konstruktiver Abschnitt

In dem konstruktiven Abschnitt verwenden wir die Eigenschaft, daß X'' ein Duplikationselement ist, um \mathcal{R}'' aus \mathcal{R} zu konstruieren. Die Idee der Konstruktion von \mathcal{R}'' ist es, die Duplikation an dem Duplikationselement X'' "tiefer" (größerer Abstand zur Wurzel) im Reconciled Tree \mathcal{R} stattfinden zu lassen. Dies geschieht durch zwei Ersetzungsschritte der Operation Repl , welche in Abbildung 4.5 durch \mathcal{R}' und \mathcal{R}'' wiedergegeben sind.

1. Wir ersetzen $T_{\mathcal{R}}(X'')$ durch $T_S(B(X'')) \times p$ mit einem zu \mathcal{R} tdisjunkten $p \in \mathbb{N}$ zu $\mathcal{R}' = \text{Repl}(\mathcal{R}, T_S(B(X'')) \times p, X'')$. Die Funktion Υ in Ab-

bildung 4.5 bezeichnet die durch die Ersetzung \mathcal{R}' gegebene Funktion $\Upsilon_{\mathcal{R}, T_{\mathcal{S}}(B(X'')) \times p, X''}$.

$T_{\mathcal{S}}(B(X'')) \times p$ ist das kleinstmögliche m -Clustersystem, welches wir ersetzen können, so daß die Ersetzung \mathcal{R}' noch spezieskonform zu \mathcal{S} ist. Dies bedeutet, daß wir alle Duplikationen aus $T_{\mathcal{R}}(X'')$ entfernen. Damit können wir möglicherweise aber auch keinen Einbettungsmonomorphismus mehr finden, um $T_{\mathcal{G}}(X)$ in \mathcal{R}' einzubetten. Dies ist dann der Fall, wenn $T_{\mathcal{R}}(X'')$ ein notwendiges Duplikationselement beinhaltet, um \mathcal{G} in \mathcal{R} spezieskonform mit φ einzubetten. Die spezieskonforme Einbettbarkeit sichern wir durch die zweite Ersetzung.

2. Wir konstruieren aus \mathcal{R}' eine Ersetzung $\mathcal{R}'' = \text{Repl}(\mathcal{R}', \mathcal{D}, \bigsqcup_{\mathcal{S}}(X) \times \{q\})$, in welche \mathcal{G} wieder spezieskonform einbettbar ist. Die in Abbildung 4.5 gegebene Funktion Υ' bezeichnet die durch die Ersetzung $T_{\mathcal{R}}''$ gegebene Funktion $\Upsilon_{\mathcal{R}', \mathcal{D}, \bigsqcup_{\mathcal{S}}(X) \times \{q\}}$.

$\overline{T_{\mathcal{G}}}(X)$ ist in $\overline{T_{\mathcal{R}'}}(X)$ spezieskonform einbettbar, da wir hier keine Ersetzung durchgeführt haben. Die zu $\overline{T_{\mathcal{G}}}(X)$ in $T_{\mathcal{G}}$ komplementäre Menge $T_{\mathcal{G}}(X)$ ist aber möglicherweise nicht mehr spezieskonform in \mathcal{R}' einbettbar. Daher konstruieren wir mit Merge ein m -Clustersystem \mathcal{D} , in welches wir $T_{\mathcal{G}}(X)$ spezieskonform einbetten können und ersetzen dieses in \mathcal{R}' an Element $\bigsqcup_{\mathcal{S}}(X) \times \{q\}$.

Alle Elemente des Teilbaumes $T_{\mathcal{G}}(X)$ ohne X werden in \mathcal{R} spezieskonform in die Teilbäume $T_{\mathcal{R}}(Z'_1), \dots, T_{\mathcal{R}}(Z''_n)$ eingebettet. Daher verwenden wir diese Teilbäume, um sie mit Merge unter der neuen Wurzel Y'' zu \mathcal{D} zusammenzufügen. Y'' besitzt dabei die gleiche Basis wie seine Kinder und ist damit ein Duplikationselement. Da Y'' wie X'' alle Teilbäume $T_{\mathcal{R}}(Z'_1), \dots, T_{\mathcal{R}}(Z''_n)$ beinhaltet, kann X in Y'' spezieskonform eingebettet werden. Damit übernimmt Y'' die in \mathcal{R} an dem Duplikationselement X'' stattfindende Duplikation. Somit ist $T_{\mathcal{G}}(X)$ spezieskonform in \mathcal{D} einbettbar.

Abschließend ersetzen wir \mathcal{D} in \mathcal{R}' . Da $\bigsqcup_{\mathcal{S}}(X) = B(Y''_i)$ gilt und Y'' das Duplikationselement der Kopien Y''_1, \dots, Y''_n ist, folgt $B(Y'') = \bigsqcup_{\mathcal{S}}(X)$. Also können wir den Teilbaum $T_{\mathcal{R}'}(\bigsqcup_{\mathcal{S}}(X) \times \{p\})$ durch \mathcal{D} in \mathcal{R}' ersetzen und erhalten \mathcal{R}'' .

Somit ist \mathcal{G} in \mathcal{R}'' spezieskonform einbettbar.

Da wir \mathcal{R}'' nur durch die Operationen Repl und Merge mit von \mathcal{S} abgeleiteten Genbäumen konstruiert haben, ist \mathcal{R}'' nach Lemma 4.5 (Seite 48) und Lemma 4.6 (Seite 49) von \mathcal{S} ableitbar. Wie wir in Theorem 4.1 (Seite 43) und Theorem 4.2 (Seite 45) gezeigt haben, sind die Operationen Repl und Merge wohldefiniert. Also ist \mathcal{R}'' ein m -Clustersystem.

Damit $\mathcal{R}'' \in \text{RC}(\mathcal{G}, \mathcal{S})$ gilt, bleibt die spezieskonforme Einbettbarkeit von \mathcal{G} in \mathcal{R}'' und die minimale Kardinalität von \mathcal{R}'' , also $|\mathcal{R}| = |\mathcal{R}''|$, zu zeigen.

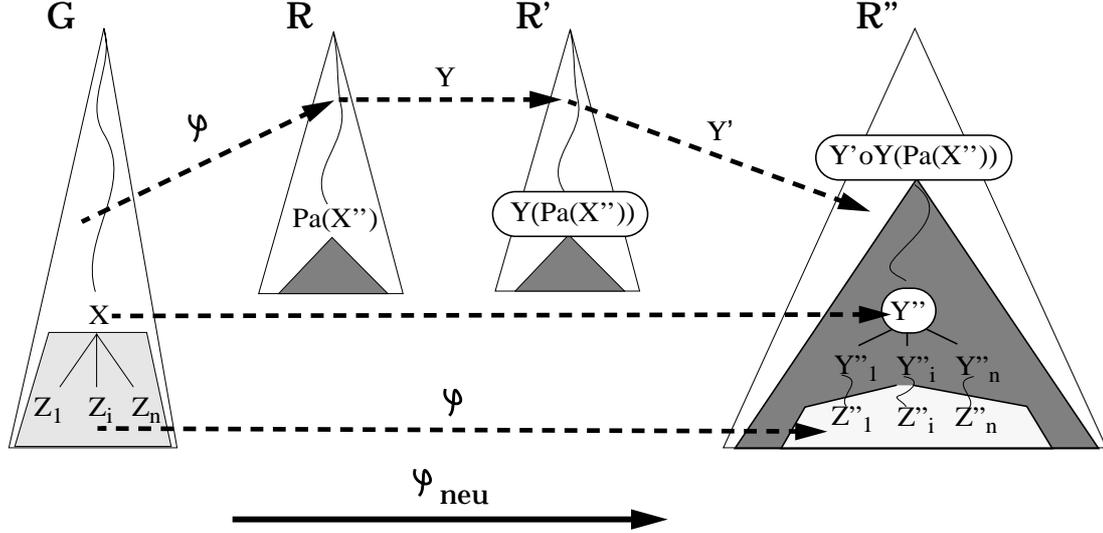


Abbildung 4.6: Der Einbettungsmonomorphismus φ_{neu} von \mathcal{G} nach \mathcal{R}''

Konstruktion des Einbettungsmonomorphismus φ_{neu} :

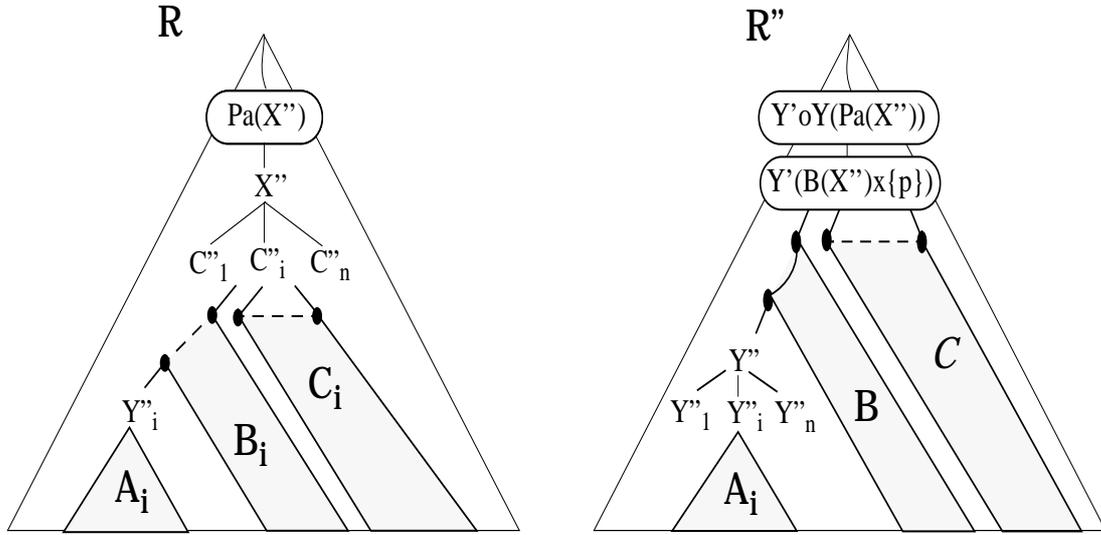
\mathcal{R}'' haben wir so konstruiert, daß \mathcal{G} dort monomorph und spezieskonform eingebettet werden kann. Abbildung 4.6 zeigt eine solche Abbildung φ_{neu} , welche aus partiellen Abbildungen von φ und mit Hilfe der Einbettungsisomorphismen Υ und Υ' konstruiert wird. Zur Konstruktion partitionieren wir \mathcal{G} in $\overline{T_{\mathcal{G}}(X)}$, $\{X\}$ und $T_{\mathcal{G}}(X) \setminus \{X\}$. Für jedes Element dieser Partition konstruieren wir nun partiell φ_{neu} .

Wir betrachten zuerst die Urbildmenge $\overline{T_{\mathcal{G}}(X)}$, welche unter φ auf $\overline{T_{\mathcal{R}}(X)}$ abbildet. Die Topologie von $\overline{T_{\mathcal{R}}(X)}$ ist gleich der von $\overline{T_{\mathcal{R}''}(X'')}$, da bei den Ersetzungen lediglich die Bezeichnungen der Elemente verändert worden sind. Die veränderten Bezeichnungen erhalten wir durch die Einbettungsisomorphismen Υ' und Υ'' . Damit bilden wir die Funktion $\varphi_{\text{neu}}|_{\overline{T_{\mathcal{R}}(X)}}: \overline{T_{\mathcal{R}}(X)} \rightarrow \overline{T_{\mathcal{R}''}(Y' \circ \Upsilon(X''))}$ mit $\varphi_{\text{neu}}(Y) \mapsto \Upsilon'' \circ \Upsilon' \circ \varphi(Y)$ ab.

Das Element X bildet in \mathcal{R} auf das Duplikationselement X'' ab. Diese Funktion übernimmt in \mathcal{R}'' das Element Y'' . Daher bilden wir $\varphi_{\text{neu}}(X)$ auf Y'' ab.

$T_{\mathcal{G}}(X) \setminus \{X\}$ bilden wir mit $\varphi_{\text{neu}}|_{T_{\mathcal{G}}(X) \setminus \{X\}} := \varphi|_{T_{\mathcal{G}}(X) \setminus \{X\}}$ nach $T_{\mathcal{R}''}(Y'') \setminus \{Y''\}$ ab, da die Teilbäume $T_{\mathcal{R}}(Z''_1), \dots, T_{\mathcal{R}}(Z''_n)$ unverändert in $T_{\mathcal{R}''}(Y'') \setminus \{Y''\}$ enthalten sind.

φ_{neu} bildet somit offensichtlich (welches formal in Theorem 4.4 gezeigt ist) einen spezieskonformen Einbettungsmonomorphismus von \mathcal{G} nach \mathcal{R}'' .

Abbildung 4.7: Zu den Kardinalitäten $|\mathcal{R}|$ und $|\mathcal{R}''|$

Minimalität von \mathcal{R}'' :

Es verbleibt für \mathcal{R}'' die Minimalitätseigenschaft zu zeigen, damit $\mathcal{R}'' \in \text{RC}(\mathcal{G}, \mathcal{S})$ gilt. Da \mathcal{R} nach Annahme ein Reconciled Tree ist, müßte $|\mathcal{R}| \leq |\mathcal{R}''|$ gelten.

Wir haben \mathcal{R}'' durch ein "Nach-unten-Verschieben" des Duplikationselementes X'' jedoch so konstruiert, daß $|\mathcal{R}| \geq |\mathcal{R}''| + 2$ gilt. Dies veranschaulichen wir uns an Abbildung 4.7. Da wir nur "unter" dem Element $\text{Pa}(X'')$ ersetzt haben, gilt $|\overline{\text{T}}_{\mathcal{R}}(X'')| = |\overline{\text{T}}_{\mathcal{R}''}(\Upsilon(B(X'') \times \{p\}))|$. Wir vergleichen somit nur noch $|\text{T}_{\mathcal{R}}(X'')|$ mit $|\text{T}_{\mathcal{R}''}(\Upsilon'(B(X'') \times \{p\}))|$. Es gilt:

$$\begin{aligned}
 |\text{T}_{\mathcal{R}}(X'')| &= |\{X''\}| \\
 &\quad + \sum_{i \in \{1, \dots, n\}} (|\mathcal{A}_i| + |\mathcal{B}_i| + |\mathcal{C}_i| + |\{C''_i\}|) \\
 |\text{T}_{\mathcal{R}''}(\Upsilon'(B(X'') \times \{p\}))| &= |\Upsilon'(B(X'') \times \{p\})| \\
 &\quad + \left(\sum_{i \in \{1, \dots, n\}} |\mathcal{A}_i| \right) + |\mathcal{B}| + |\mathcal{C}| + |\{Y''\}|.
 \end{aligned}$$

Entsprechend unserer Konstruktion enthält $\text{T}_{\mathcal{R}''}(\Upsilon'(B(X'') \times \{p\}))$ alle Teilbäume $\text{T}_{\mathcal{R}}(Y''_i)$ mit $i \in \{1, \dots, n\}$, welche in der Abbildung mit \mathcal{A}_i bezeichnet sind. Wir vergleichen also noch

$$\sum_{i \in \{1, \dots, n\}} (|\mathcal{B}_i| + |\mathcal{C}_i| + 1) \text{ mit } |\mathcal{B}| + |\mathcal{C}| + 1.$$

\mathcal{B} und \mathcal{C} stammen aus der ersten Ersetzung \mathcal{R}' . Bei dieser Ersetzung wurde $\text{T}_{\mathcal{R}}(X'')$ in \mathcal{R} durch $B(X'') \times p$ ersetzt. Diese Ersetzung entspricht einer Auf-

hebung aller Duplikationen in $T_{\mathcal{R}}(X'')$. Hieraus folgert zum einen, daß alle Duplikate \mathcal{B}_i und \mathcal{C}_i durch \mathcal{B} und \mathcal{C} ersetzt wurden. Zum anderen, daß die Menge \mathcal{B} jeder um die Duplikationen bereinigten Menge \mathcal{B}_i entspricht. Also gilt $|\mathcal{B}| \leq |\mathcal{B}_i|$, für alle $i \in \{1, \dots, n\}$. Entsprechend gilt $|\mathcal{C}| \leq |\mathcal{C}_i|$, für alle $i \in \{1, \dots, n\}$. Es folgt

$$\sum_{i \in \{1, \dots, n\}} (|\mathcal{B}_i| + |\mathcal{C}_i| + |\{C_i''\}|) \geq n(|\mathcal{B}| + |\mathcal{C}| + 1).$$

Wir vergleichen somit $n(|\mathcal{B}| + |\mathcal{C}| + 1)$ mit $|\mathcal{B}| + |\mathcal{C}| + 1$. Da \mathcal{R} ein m -Clustersystem ist, gilt $n \geq 2$. Die Menge \mathcal{B} kann gleich \emptyset entsprechen, wohingegen \mathcal{C} mindestens ein Element beinhalten muß. Der geringste Unterschied zwischen $|\mathcal{R}|$ und $|\mathcal{R}''|$ beträgt somit 2 (diese Schranke ist scharf).

Damit ist \mathcal{R} kein Reconciled Tree für \mathcal{G} unter \mathcal{S} , was im Widerspruch zu unserer Annahme steht.

Die Intuition der Beweisskizze werden wir in dem folgenden Theorem formal zeigen.

4.3.2 Der formale LCA-Beweis

Theorem 4.4

Sei $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$ ein Reconciled Tree von \mathcal{G} über \mathcal{S} und φ eine minimale spezieskonforme Einbettung von \mathcal{G} in \mathcal{R} . Dann gilt $B(\varphi(X)) = \sqcup_{\mathcal{S}}(X)$, für jedes $X \in \mathcal{G}$.

Beweis. (indirekt)

Annahme: Es existiert ein $W \in \mathcal{G}$ mit $\sqcup_{\mathcal{S}}(W) \neq B(\varphi(W))$.

Wir werden zeigen, daß diese Annahme im Widerspruch zu der Minimalität von \mathcal{R} steht.

Strukturaussagen:

Wir beginnen damit, die Existenz eines Elementes $X \in \mathcal{G}$ mit $\sqcup_{\mathcal{S}}(X) \neq B(\varphi(X))$ zu zeigen, für das kein $U \in \mathcal{G}$ mit $\sqcup_{\mathcal{S}}(U) \neq B(\varphi(U))$ und $U \subset X$ existiert. Wir bezeichnen diese Eigenschaft von X als *minimal bezüglich $\sqcup_{\mathcal{S}}$* .

Aussage 1 *In \mathcal{G} existiert ein Element, welches minimal bezüglich $\sqcup_{\mathcal{S}}$ ist.*

Beweis. Da nach Voraussetzung φ spezieskonform ist, gilt $W \subseteq B(\varphi(W))$. Mit Lemma 3.4 (Seite 32) folgt hieraus $\sqcup_{\mathcal{S}}(W) \subseteq B(\varphi(W))$. Aus der Annahme $\sqcup_{\mathcal{S}}(W) \neq B(\varphi(W))$ folgt damit $\sqcup_{\mathcal{S}}(W) \subset B(\varphi(W))$. Da \mathcal{G} nach Voraussetzung endlich ist, existiert in $T_{\mathcal{G}}(W)$ ein minimales Element bezüglich $\sqcup_{\mathcal{S}}$. \square

Sei $X \in \mathcal{G}$ minimal bzgl. \sqcup_s , von nun an fest gewählt.

Aussage 2 *Es gilt $X \notin \text{Le}(\mathcal{G})$.*

Beweis. (indirekt)

Annahme: $X \in \text{Le}(\mathcal{G})$. Wir werden hieraus folgern, daß φ , im Widerspruch zur Voraussetzung, nicht minimal ist. Sei $X'' := \varphi(X)$. Aus $X \in \text{Le}(X)$ folgt $X = \sqcup_s(X)$, und da X minimal bzgl. \sqcup_s ist, gilt $\sqcup_s(X) \subset B(X'')$, woraus $X \subset B(X'')$ folgt.

Da X ein Blatt ist, existiert nach Def. 2.4 (Seite 17) ein $a \in M$ mit $\{a\} = X$. Also gilt $\{a\} \subset B(X'')$, woraus die Existenz eines $i \in \mathbb{N}$ mit $(a, i) \in X''$ folgt. Aus Def. 2.4 (Seite 17) folgt $\{(a, i)\} \in \mathcal{R}$. Da $\{a\} \subset B(X'')$ gilt, folgt weiter $\{(a, i)\} \subset X''$. Also existiert ein $\{(a, i)\} \in \mathcal{R}$ mit $\{(a, i)\} \subset X''$ und $X = \{a\} = B(\{(a, i)\})$.

Wir konstruieren nun, mit Hilfe des Elementes $\{(a, i)\}$, aus φ eine minimalere Einbettung φ_{neu} , entsprechend Def. 3.6 (Seite 3.6).

$$\mathcal{V} := \{F \in \text{Emon}(\mathcal{G}, \mathcal{R}) \mid \forall V \in \mathcal{G} : V \subseteq B(F(V))\}$$

ist die Menge aller spezieskonformen Einbettungsmonomorphismen von \mathcal{G} nach \mathcal{R} . Nach Voraussetzung ist φ ein minimales Element von \mathcal{V} . Da $\varphi \in \text{Emon}(\mathcal{G}, \mathcal{R})$ und $X \in \text{Le}(\mathcal{G})$ ist, folgt $\varphi(\mathcal{G}) \cap (\mathcal{T}_{\mathcal{R}}(X'') \setminus X'') = \emptyset$. Also ist $\varphi_{\text{neu}} : \mathcal{G} \rightarrow \mathcal{R}$, mit

$$\varphi_{\text{neu}}(V) := \begin{cases} \varphi(V) & V \neq \{a\} \\ \{(a, i)\} & V = \{a\} \end{cases}$$

ein Element aus \mathcal{V} . Es gilt somit $\varphi_{\text{neu}} < \varphi$, welches ein Widerspruch zu der vorausgesetzten Minimalität von φ ist. \square

Wir bezeichnen die Kinder von X durch Z_1, \dots, Z_n , mit $n \in \mathbb{N}$. Die folgende Aussage zeigt Struktureigenschaften von \mathcal{R} auf, welche uns die Anwendung der Operationen aus Abschnitt 4.1 ermöglichen. Zur Vereinfachung definieren wir $X'' := \varphi(X)$ und $Z_i'' := \varphi(Z_i)$ für alle $i \in \{1, \dots, n\}$.

Aussage 3 *Sei $Z'' \in \varphi(\text{Ch}_{\mathcal{G}}(X))$, dann existiert ein $Y'' \in K_{\mathcal{R}}(X'', Z'')$ mit $B(Y'') = \sqcup_s(X)$ und $B(Y'') \subset B(X'')$.*

Beweis. Da $Z'' \in \varphi(\text{Ch}_{\mathcal{G}}(X))$, existiert ein $Z \in \text{Ch}_{\mathcal{G}}(X)$ mit $\varphi(Z) = Z''$.

Wir zeigen zuerst, es existiert ein $Y' \in K_s(B(X''), B(Z''))$ mit $Y' \subset B(X'')$ und $Y' = \sqcup_s(X)$. Da X minimal bzgl. \sqcup_s ist, gilt $\sqcup_s(X) \subset B(X'')$ und $\sqcup_s(Z) = B(Z'')$. Aus $Z \in \text{Ch}_{\mathcal{G}}(X)$ folgt $\sqcup_s(Z) \subseteq \sqcup_s(X)$. Somit gilt

$$B(Z'') = \sqcup_s(Z) \subseteq \sqcup_s(X) \subset B(X''), \text{ also } Y' \in K_s(B(X''), B(Z'')).$$

Wir zeigen nun die Existenz von Y'' . Da \mathcal{R} von \mathcal{S} ableitbar ist, gilt $B(K_{\mathcal{R}}(X'', Z'')) = K_{\mathcal{S}}(B(X''), B(Z''))$. Da $Y' \in K_{\mathcal{S}}(B(X''), B(Z''))$ ist, existiert somit ein

$$Y'' \in K_{\mathcal{R}}(X'', Z''), \text{ mit } B(Y'') = Y'.$$

Es gilt $Y' \subset B(X'')$, also $B(Y'') \subset B(X'')$. Aus $B(Y'') \subset B(X'')$ und $Y'' \in K_{\mathcal{R}}(X'', Z'')$ folgt $Y'' \subset X''$. Es gilt $Y' = \bigsqcup_{\mathcal{S}}(X)$, also $B(Y'') = \bigsqcup_{\mathcal{S}}(X)$. \square

Wir bezeichnen Y_i'' als das jeweils maximale Element einer Kette $K_{\mathcal{R}}(Z_i'', X'')$ mit $B(Y_i'') = \bigsqcup_{\mathcal{S}}(X)$ und $B(Y_i'') \subset B(X'')$, sowie $C_i'' \in \text{Ch}_{\mathcal{R}}(X'')$ mit $Z_i'' \subseteq C_i''$, für alle $i \in \{1, \dots, n\}$.

Aussage 4 X'' ist ein Duplikationselement.

Beweis. Da \mathcal{R} ein m -Clustersystem ist, existieren Y_i'', Y_j'' mit $i, j \in \{1, \dots, n\}$ und $i \neq j$. Aus $Y_i'' \in K_{\mathcal{S}}(X'', Z_i'')$ und $Y_i'' \subset X''$ folgt $Y_i'' \in K_{\mathcal{S}}(C_i'', Z_i'')$, entsprechend folgt $Y_j'' \in K_{\mathcal{S}}(C_j'', Z_j'')$. Damit gilt $B(Y_i'') \subseteq B(C_i'')$ und $B(Y_j'') \subseteq B(C_j'')$. Es gilt $B(Y_i'') = \bigsqcup_{\mathcal{S}}(X) = B(Y_j'')$, womit $B(C_i'') \cap B(C_j'') \neq \emptyset$ folgt und somit X'' (Def. 3.3, Seite 31) ein Duplikationselement ist. \square

Aussage 5 Sei $i \in \{1, \dots, n\}$, dann gilt $Y_i'' \subset C_i''$.

Beweis. Da $B(Y_i'') \subset B(X'')$ nach Voraussetzung gilt und $B(C_i'') = B(X'')$ nach Aussage 4 gilt, folgt $B(Y_i'') \subset B(C_i'')$. Mit der Voraussetzung $Y_i'' \in K_{\mathcal{R}}(X'', Z_i'')$ folgt $Y_i'' \subset C_i''$. \square

Konstruktion von \mathcal{R}'' :

Wir werden nun \mathcal{R}'' definieren. Die Idee der Ersetzung ist es, die Duplikation an dem Duplikationselement X'' echt unterhalb von X'' stattfinden zu lassen. Hierzu eliminieren wir zuerst die Duplikation, indem wir den duplikationsfreien Teilbaum $B(X'') \rtimes p$ mit einem zu $\text{Ro}(\mathcal{R})$ tdisjunkten $p \in (\mathbb{N})$, an X'' ersetzen.

Festlegung 1 (\mathcal{R}')

$$\mathcal{R}' := \text{Repl}(\mathcal{R}, T_{\mathcal{S}}(B(X'')) \rtimes p, X'')$$

Aussage 6 \mathcal{R}' ist eine Ersetzung.

Beweis. \mathcal{R} ist nach Voraussetzung und $T_s(B(X'')) \times p$) ist nach Theorem 4.3 (Seite 47) ein m -Clustersystem. Da p tdisjunkt zu $\text{Ro}(\mathcal{R})$ ist, gilt $\text{Ro}(\mathcal{R}) \cap \text{Ro}(T_s(B(X'')) \times p) = \emptyset$. Es gilt $B(\text{Ro}(T_s(B(X'')) \times p)) = B(B(X'') \times \{p\}) = B(X'')$ und $X'' = \varphi(X) \in \mathcal{R}$. Damit ist \mathcal{R}' eine Ersetzung. \square

Als nächstes konstruieren wir uns einen Baum, welcher als Wurzel ein Duplikationselement und die Teilbäume $T_{\mathcal{R}}(Z_i'')$ für alle $i \in \{1, \dots, n\}$ enthält.

Festlegung 2 (\mathcal{D})

$$\mathcal{D} := \text{Merge}(T_{\mathcal{R}}(Y_1''), \dots, T_{\mathcal{R}}(Y_n''))$$

Aussage 7 \mathcal{D} ist eine Zusammenfügung.

Beweis. Seien $i, j \in \{1, \dots, n\}$ mit $i \neq j$ fest aber beliebig gewählt. Wir zeigen zuerst $Y_i'' \cap Y_j'' = \emptyset$. Nach Voraussetzung gilt $Y_i'' \in K_{\mathcal{R}}(X'', Z_i'')$ und $Y_j'' \in K_{\mathcal{R}}(X'', Z_j'')$. Da $\varphi \in \text{Emon}(\mathcal{G}, \mathcal{R})$ folgt $Z_i'' \cap Z_j'' = \emptyset$ und damit gilt auch $Y_i'' \cap Y_j'' = \emptyset$.

Nach Voraussetzung gilt $B(Y_i'') = \bigsqcup_s(X)$, für alle $i \in \{1, \dots, n\}$. Damit ist \mathcal{D} eine Zusammenfügung. \square

Wir ersetzen nun \mathcal{D} an $\bigsqcup_s(X) \times \{p\}$ in \mathcal{R}' .

Festlegung 3 (\mathcal{R}'')

$$\mathcal{R}'' := \text{Repl}\left(\mathcal{R}', \mathcal{D}, \bigsqcup_s(X) \times \{p\}\right)$$

Aussage 8 \mathcal{R}'' ist die Ersetzung von \mathcal{D} an $\bigsqcup_s(X) \times \{p\}$ in \mathcal{R}' .

Beweis. Wir zeigen die Bedingungen an die Parameter:

1. $\mathcal{R}', \mathcal{D}$ sind m -Clustersysteme: Nach Aussage 6 ist \mathcal{R}' eine Ersetzung und damit nach Theorem 4.1 ein m -Clustersystem. Nach Aussage 7 ist \mathcal{D} eine Zusammenfügung und damit nach Theorem 4.2 ein m -Clustersystem.
2. $\text{Ro}(\mathcal{D}) \cap \text{Ro}(\mathcal{R}') = \emptyset$: Es gilt $\text{Ro}(\mathcal{R}') = \text{Ro}(\mathcal{R}) \setminus X'' \cup \bigsqcup_s(X) \times \{p\}$ und $\text{Ro}(\mathcal{D}) = \bigcup_{i \in \{1, \dots, n\}} Y_i''$. Da p tdisjunkt zu $\text{Ro}(\mathcal{R})$ ist, folgt $\text{Ro}(\mathcal{D}) \cap \bigsqcup_s(X) \times \{p\} = \emptyset$. Es gilt nach Voraussetzung $Y_i'' \subset X''$ für alle $i \in \{1, \dots, n\}$, woraus $\text{Ro}(\mathcal{D}) \cap \text{Ro}(\mathcal{R}) \setminus X'' = \emptyset$ folgt. Somit gilt die Aussage.
3. $\bigsqcup_s(X) \times \{p\} \in \mathcal{R}'$. Nach Voraussetzung gilt $\bigsqcup_s(X) \neq B(X'')$, und mit Lemma 3.4 folgt $\bigsqcup_s(X) \subset B(X'')$. Hieraus folgt $\bigsqcup_s(X) \in T_s(B(X''))$, und damit gilt $\bigsqcup_s(X) \times \{p\} \in T_s(B(X'')) \times p$. Da $T_s(B(X'')) \times p \subseteq \mathcal{R}'$ ist, folgt die Aussage.

4. $B(\text{Ro}(\mathcal{D})) = B(\bigsqcup_{\mathcal{S}}(X) \times \{p\})$: Es gilt $\text{Ro}(\mathcal{D}) = \bigcup_{i \in \{1, \dots, n\}} Y_i''$ und $B(Y_i'') = \bigsqcup_{\mathcal{S}}(X)$ für alle $i \in \{1, \dots, n\}$. Also ist

$$B(\text{Ro}(\mathcal{D})) = B\left(\bigcup_{i \in \{1, \dots, n\}} Y_i''\right) = \bigsqcup_{\mathcal{S}}(X) = B\left(\bigsqcup_{\mathcal{S}}(X) \times \{p\}\right).$$

□

Durch die Ersetzungen \mathcal{R}' und \mathcal{R}'' existieren jeweils die Funktionen

$$\begin{aligned} \Upsilon_{\mathcal{R}, \bigsqcup_{\mathcal{S}}(X) \times p, X''} &: \overline{T}_{\mathcal{R}}(X'') \rightarrow \overline{T}_{\mathcal{R}'}(B(X'') \times \{p\}) \\ \Upsilon_{\mathcal{R}', \mathcal{D}, \bigsqcup_{\mathcal{S}}(X) \times \{p\}} &: \overline{T}_{\mathcal{R}'}\left(\bigsqcup_{\mathcal{S}}(X) \times \{p\}\right) \rightarrow \overline{T}_{\mathcal{R}''}(\text{Ro}(\mathcal{D})) \end{aligned}$$

Zur Vereinfachung verwenden wir für diese Funktionen die Kurzbezeichnungen,

$$\Upsilon' := \Upsilon_{\mathcal{R}, \bigsqcup_{\mathcal{S}}(X) \times p, X''} \quad \text{und} \quad \Upsilon'' := \Upsilon_{\mathcal{R}', \mathcal{D}, \bigsqcup_{\mathcal{S}}(X) \times \{p\}}.$$

Die Funktionen Υ' und Υ'' werden wir nun verwenden, um mit Hilfe von φ einen Einbettungsmonomorphismus von \mathcal{G} nach \mathcal{R}'' zu konstruieren. Hierzu partitionieren wir \mathcal{G} in $\overline{T}_{\mathcal{G}}(X)$, $\{X\}$ und $T_{\mathcal{G}}(X) \setminus \{X\}$. Aus $\overline{T}_{\mathcal{G}}(X)$ bilden wir mit $\Upsilon'' \circ \Upsilon' \circ \varphi|_{\overline{T}_{\mathcal{G}}(X)}$ nach \mathcal{R}'' ab.

Aussage 9 *Es gilt $\Upsilon'' \circ \Upsilon' \circ \varphi|_{\overline{T}_{\mathcal{G}}(X)} \in \text{Emon}(\overline{T}_{\mathcal{G}}(X), \overline{T}_{\mathcal{R}''}(I))$ mit $I := B(X'') \times \{p\} \setminus \bigsqcup_{\mathcal{S}}(X) \times \{p\} \cup \text{Ro}(\mathcal{D})$.*

Beweis. Wir zeigen zuerst $\Upsilon'' \circ \Upsilon' \in \text{Emon}(\overline{T}_{\mathcal{R}}(X''), \overline{T}_{\mathcal{R}''}(\Upsilon''(B(X'') \times \{p\})))$. Der Wertebereich von Υ' ist $\overline{T}_{\mathcal{R}''}(B(X'') \times \{p\})$, und der Definitionsbereich von Υ'' ist $\overline{T}_{\mathcal{R}''}(\bigsqcup_{\mathcal{S}}(X) \times \{p\})$. Da X minimal bzgl. $\bigsqcup_{\mathcal{S}}$ ist, gilt $\bigsqcup_{\mathcal{S}}(X) \subset B(X'')$ und damit $\bigsqcup_{\mathcal{S}}(X) \times \{p\} \subset B(X'') \times \{p\}$. Also ist $\Upsilon'' \circ \Upsilon'$ eine Funktion. Da Υ' und Υ'' Ordnungsisomorphismen (Lemma 4.2) sind, ist $\Upsilon'' \circ \Upsilon' \in \text{Oiso}(\overline{T}_{\mathcal{R}}(X''), I)$, mit $I = \overline{T}_{\mathcal{R}''}(\Upsilon''(B(X'') \times \{p\})) = \overline{T}_{\mathcal{R}''}(B(X'') \times \{p\} \setminus \bigsqcup_{\mathcal{S}}(X) \times \{p\} \cup \text{Ro}(\mathcal{D}))$.

Die eigentliche Aussage folgt nun direkt aus der Voraussetzung

$$\varphi|_{\overline{T}_{\mathcal{G}}(X)} \in \text{Emon}(\overline{T}_{\mathcal{G}}(X), \overline{T}_{\mathcal{R}}(X'')).$$

□

Aus $T_{\mathcal{G}}(X) \setminus \{X\}$ können wir direkt mit $\varphi|_{T_{\mathcal{G}}(X) \setminus \{X\}}$ nach \mathcal{R}'' abbilden. Denn \mathcal{R}'' beinhaltet \mathcal{D} als Teilbaum. \mathcal{D} wiederum beinhaltet alle Mengen aus $T_{\mathcal{R}}$, auf welche $\varphi|_{T_{\mathcal{G}}(X) \setminus \{X\}}$ abbildet. Um die Einbettungsmonomorphie zu erhalten, bilden wir X auf $\text{Ro}(\mathcal{D})$ ab. Wir konstruieren unseren neuen Einbettungsmonomorphismus φ_{neu} somit wie folgt.

Festlegung 4 $\varphi_{\text{neu}} : \mathcal{G} \rightarrow \mathcal{R}''$ ist wie folgt definiert:

$$\varphi_{\text{neu}}(K) := \begin{cases} \Upsilon'' \circ \Upsilon' \circ \varphi|_{\overline{T_{\mathcal{G}}(X)}} & K \in \overline{T_{\mathcal{G}}(X)} \\ \text{Ro}(\mathcal{D}) & K = X \\ \varphi|_{T_{\mathcal{G}}(X) \setminus \{X\}} & \text{sonst.} \end{cases}$$

Aussage 10 Es gilt $\varphi_{\text{neu}} \in \text{Emon}(\mathcal{G}, \mathcal{R}'')$.

Beweis. Wir zeigen $\varphi_{\text{neu}}(A \sqcup B) = \varphi_{\text{neu}}(A) \sqcup \varphi_{\text{neu}}(B)$ o.B.d.A. durch die folgenden Fälle:

1. $A, B \in \overline{T_{\mathcal{G}}(X)}$: gilt nach Aussage 9.
2. $A, B \in T_{\mathcal{G}}(Z_i)$ für ein festes aber beliebiges $i \in \{1, \dots, n\}$: $T_{\mathcal{G}}(Z_i)$ ist unter \sqcup abgeschlossen und $\varphi \in \text{Emon}(\mathcal{G}, \mathcal{R})$. Damit folgt $\varphi|_{T_{\mathcal{G}}(Z_i)} \in \text{Emon}(T_{\mathcal{G}}(Z_i), T_{\mathcal{R}}(Z_i''))$. Nach Voraussetzung gilt $\varphi_{\text{neu}}|_{T_{\mathcal{G}}(Z_i)} = \varphi|_{T_{\mathcal{G}}(Z_i)}$ und somit $\varphi_{\text{neu}}|_{T_{\mathcal{G}}(Z_i)} \in \text{Emon}(T_{\mathcal{G}}(Z_i), T_{\mathcal{R}}(Z_i''))$.
3. $A \in T_{\mathcal{G}}(Z_i), B \in T_{\mathcal{G}}(Z_j)$ mit festen aber beliebigen $i, j \in \{1, \dots, n\}$ und $i \neq j$:

$$\varphi_{\text{neu}}(A \sqcup_{\mathcal{G}} B) = \varphi_{\text{neu}}(X) \tag{4.7}$$

$$\begin{aligned} &= \text{Ro}(\mathcal{D}) \\ &= Y_i'' \sqcup Y_j'' = \text{Ro}(\mathcal{D}) \end{aligned} \tag{4.8}$$

$$\begin{aligned} &= \min(K_{\mathcal{D}}(\text{Ro}(\mathcal{D}), Y_i'') \cap K_{\mathcal{D}}(\text{Ro}(\mathcal{D}), Y_j'')) \\ &= \min(K_{\mathcal{D}}(\text{Ro}(\mathcal{D}), Z_i'') \cap K_{\mathcal{D}}(\text{Ro}(\mathcal{D}), Z_j'')) \end{aligned} \tag{4.9}$$

$$\begin{aligned} &= \min(K_{\mathcal{R}''}(\text{Ro}(\mathcal{D}), Z_i'') \cap K_{\mathcal{R}''}(\text{Ro}(\mathcal{D}), Z_j'')) \\ &= \min(K_{\mathcal{R}''}(\text{Ro}(\mathcal{R}''), Z_i'') \cap K_{\mathcal{R}''}(\text{Ro}(\mathcal{R}''), Z_j'')) \\ &= \min(K_{\mathcal{R}''}(\text{Ro}(\mathcal{R}''), \varphi_{\text{neu}}(A)) \cap K_{\mathcal{R}''}(\text{Ro}(\mathcal{R}''), \varphi_{\text{neu}}(B))) \\ &= \varphi_{\text{neu}}(A) \sqcup_{\mathcal{R}''} \varphi_{\text{neu}}(B) \end{aligned}$$

Zu (4.7): Es gilt

$$\begin{aligned} A \sqcup_{\mathcal{G}} B &= \min(K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), A) \cap K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), B)) \\ &= \min(K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), Z_i) \cap K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), Z_j)) \\ &= X \end{aligned}$$

Zu (4.8): Nach Definition der Zusammenfügung gilt $Y_i'', Y_j'' \in \text{Ch}_{\mathcal{D}}(\text{Ro}(\mathcal{D}))$, also $Y_i'' \sqcup Y_j'' = \text{Ro}(\mathcal{D})$.

Zu (4.9): Nach Voraussetzung der Zusammenfügung \mathcal{D} gilt $Y_i'' \cap Y_j'' = \emptyset$. Nach Voraussetzung von Y_i'' gilt $Z_i'' \subseteq Y_i''$ und von Y_j'' gilt $Z_j'' \subseteq Y_j''$. Damit folgt $K_{\mathcal{D}}(Y_i'', Z_i'') \cap K_{\mathcal{D}}(Y_j'', Z_j'') = \emptyset$.

4. $A = X, B \in T_{\mathcal{G}}(X)$: Also gilt $B \subseteq A$ und damit $B \sqcup_{\mathcal{G}} A = A$. Wenn $B = X$ gilt, ist $\varphi_{\text{neu}}(B) = \varphi_{\text{neu}}(A)$. Sonst existiert ein $i \in \{1, \dots, n\}$ mit $B \subseteq T_{\mathcal{G}}(Z_i)$. Damit folgt $\varphi_{\text{neu}}(B) \subseteq Z_i'' \subseteq Y_i'' \subset \text{Ro}(\mathcal{D}) = \varphi_{\text{neu}}(X)$. Also gilt immer $\varphi_{\text{neu}}(B) \subseteq \varphi_{\text{neu}}(X)$, woraus $\varphi_{\text{neu}}(A) \sqcup_{\mathcal{R}''} \varphi_{\text{neu}}(B) = \varphi_{\text{neu}}(A)$ folgt.

Somit gilt, $\varphi_{\text{neu}}(A \sqcup_{\mathcal{G}} B) = \varphi_{\text{neu}}(A) = \varphi_{\text{neu}}(B) \sqcup_{\mathcal{R}''} \varphi_{\text{neu}}(A) = \varphi_{\text{neu}}(A)$.

5. $A \in \overline{T_{\mathcal{G}}}(X), B \in T_{\mathcal{G}}(X)$: Hieraus folgt $A \sqcup B = A \sqcup \text{Pa}(X)$. Da $\text{Pa}(X), A \in \overline{T_{\mathcal{G}}}(X)$ sind, lassen sich diese beiden Elemente einbettungsmonomorph (siehe Aussage 9) nach \mathcal{R}'' abbilden. Damit folgt

$$\varphi_{\text{neu}}(A \sqcup B) = \varphi_{\text{neu}}(A \sqcup \text{Pa}(X)) = \varphi_{\text{neu}}(A) \sqcup \varphi_{\text{neu}}(\text{Pa}(X)).$$

Wir zeigen nun, daß $\varphi_{\text{neu}}(A) \sqcup \varphi_{\text{neu}}(\text{Pa}(X)) = \varphi_{\text{neu}}(A) \sqcup \varphi_{\text{neu}}(B)$ gilt, indem wir die Existenz eines $C'' \in \text{Ch}(\varphi_{\text{neu}}(\text{Pa}(X)))$ mit $\varphi_{\text{neu}}(A) \in \overline{T_{\mathcal{R}''}}(C'')$ und $\varphi_{\text{neu}}(B) \in T_{\mathcal{R}''}''(C'')$ zeigen.

Da $\varphi \in \text{Emon}(\mathcal{G}, \mathcal{R})$ ist, folgt $\varphi(X) \subset \varphi(\text{Pa}(X))$. Damit existiert ein $C \in \text{Ch}(\varphi(\text{Pa}(X)))$ mit $\varphi(X) \subseteq C$. Aus C folgern wir nun die Existenz von C'' , wobei wir die beiden folgenden Fälle zu unterscheiden haben.

- (a) $X'' \subset C$: Damit ist $C \in \overline{T_{\mathcal{R}}}(X'')$, und wir können C mit $\Upsilon'' \circ \Upsilon'$ nach $T_{\mathcal{R}}''$ abbilden. Es gilt $\Upsilon' \in \text{Eiso}(\overline{T_{\mathcal{R}}}(X''), \overline{T_{\mathcal{R}'}}(B(X'') \times \{p\}))$ und $\Upsilon'' \in \text{Eiso}(\overline{T_{\mathcal{R}'}}(\bigsqcup_{\mathcal{S}}(X) \times \{p\}), \overline{T_{\mathcal{R}''}}(\text{Ro}(\mathcal{D})))$. Damit folgt aus $C \in \text{Ch}(\varphi(\text{Pa}(X)))$, daß $\Upsilon'' \circ \Upsilon'(C) \in \text{Ch}(\Upsilon'' \circ \Upsilon'(\varphi(\text{Pa}(X))))$ gilt. Also ist $C'' \in \text{Ch}(\varphi_{\text{neu}}(\text{Pa}(X)))$ mit $C'' := \Upsilon'' \circ \Upsilon'(C)$.

Wir zeigen $\varphi_{\text{neu}}(A) \in \overline{T_{\mathcal{R}''}}(C'')$. Da φ ein Einbettungsmonomorphismus und $A \in \overline{T_{\mathcal{G}}}(X)$ ist, folgt $\varphi(A) \in \overline{T_{\mathcal{R}}}(C)$. $\Upsilon'' \circ \Upsilon'$ ist nach Aussage 9 ein Einbettungsisomorphismus, womit $\Upsilon'' \circ \Upsilon' \circ \varphi(A) \in \overline{T_{\mathcal{R}''}}(\Upsilon'' \circ \Upsilon'(C))$ ist. Also, ist $\varphi_{\text{neu}}(A) \in \overline{T_{\mathcal{R}''}}(C'')$.

Wir zeigen $\varphi(B) \in T_{\mathcal{R}''}''(C'')$. Da $T_{\mathcal{R}}(X'')$ durch $T_{\mathcal{S}}(B(X'')) \times p$ in \mathcal{R}' ersetzt wird und $X'' \subset C$ ist, folgt $\Upsilon'(C) = C \setminus X'' \cup B(X'') \times \{p\}$. Damit ist $B(X'') \times \{p\} \subseteq \Upsilon'(C)$. $T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \times p$ wird in \mathcal{R}'' durch \mathcal{D} ersetzt. Es gilt $\bigsqcup_{\mathcal{S}}(X) \subset B(X'')$, woraus $\bigsqcup_{\mathcal{S}}(X) \times \{p\} \subset B(X'') \times \{p\}$ folgt, und es gilt $B(X'') \times \{p\} \subseteq \Upsilon'(C)$. Damit ist $\Upsilon'' \circ \Upsilon'(C) = \Upsilon'(C) \setminus \bigsqcup_{\mathcal{S}}(X) \times \{p\} \cup \mathcal{D}$. Da $\varphi(B) \in \mathcal{D}$ gilt, folgt $\varphi(B) \in T_{\mathcal{R}''}''(C'')$.

- (b) $X'' = C$: Damit wird $T_{\mathcal{R}}(C)$ durch $T_{\mathcal{S}}(B(X'')) \times p$ in \mathcal{R}'' ersetzt, womit $B(X'') \times \{p\} \in \text{Ch}(\Upsilon' \circ \varphi(\text{Pa}(X)))$ ist. Da $\bigsqcup_{\mathcal{S}}(X) \times \{p\} \subset B(X'') \times \{p\}$ gilt, sind $B(X'') \times \{p\}$ und $\Upsilon' \circ \varphi(\text{Pa}(X))$ mit $\Upsilon'' \in \text{Eiso}(\overline{T_{\mathcal{R}'}}(\bigsqcup_{\mathcal{S}}(X) \times \{p\}), \overline{T_{\mathcal{R}''}}(\text{Ro}(\mathcal{D})))$ einbettungsisomorph auf \mathcal{R}'' abbildbar. Damit folgt $\Upsilon''(B(X'') \times \{p\}) \in \text{Ch}(\Upsilon'' \circ \Upsilon' \circ \varphi(\text{Pa}(X)))$ aus

$B(X'') \times \{p\} \in \text{Ch}(\Upsilon' \circ \varphi(\text{Pa}(X)))$. Also ist $C'' \in \text{Ch}(\varphi_{\text{neu}}(\text{Pa}(X)))$ mit $C'' := \Upsilon''(B(X'') \times \{p\})$.

Wir zeigen $\varphi_{\text{neu}}(A) \in \overline{T_{\mathcal{R}''}}(C'')$. Da $\sqcup_S(X) \subset B(X'')$ gilt, folgt $\sqcup_S(X) \times \{p\} \subset B(X'') \times \{p\}$. $T_S(\sqcup_S(X)) \times p$ wird in \mathcal{R}'' durch \mathcal{D} ersetzt. Also ist $C'' = B(X'') \times \{p\} \setminus \sqcup_S(X) \times \{p\} \cup \mathcal{D}$. Aus Aussage 9 und $A \in \overline{T_{\mathcal{G}}}(X)$ folgt $\varphi_{\text{neu}}(A) \in \overline{T_{\mathcal{R}''}}(C'')$.

Wir zeigen $\varphi(B) \in T_{\mathcal{R}''}(C'')$. Da $C'' = B(X'') \times \{p\} \setminus \sqcup_S(X) \times \{p\} \cup \mathcal{D}$ ist, folgt $\text{Ro}(\mathcal{D}) \subseteq C''$. Mit $\varphi_{\text{neu}}(B) \in \mathcal{D}$ folgt $\varphi_{\text{neu}}(B) \in T_{\mathcal{R}''}(C'')$.

□

Aussage 11 \mathcal{R}'' ist von \mathcal{S} ableitbar.

Beweis. Es gilt $\mathcal{R}' = \text{Repl}(\mathcal{R}, T_S(B(X'')) \times p, X'')$. Da \mathcal{R} ein Reconciled Tree ist, ist dieser von \mathcal{S} ableitbar. $T_S(B(X'')) \times p$ ist offensichtlich von \mathcal{S} ableitbar. Aus Lemma 4.5 folgt damit, daß \mathcal{R}' aus \mathcal{S} ableitbar ist.

Es gilt $\mathcal{R}'' = \text{Repl}(\mathcal{R}', \mathcal{D}, \sqcup_S(X'') \times \{p\})$ und $\mathcal{D} = \text{Merge}(T_{\mathcal{R}}(Y_1''), \dots, T_{\mathcal{R}}(Y_n''))$. Da \mathcal{R} von \mathcal{S} ableitbar ist, sind auch dessen Teilbäume $T_{\mathcal{R}}(Y_1''), \dots, T_{\mathcal{R}}(Y_n'')$ von \mathcal{S} ableitbar. Nach Lemma 4.6 ist \mathcal{D} damit von \mathcal{S} ableitbar. Da \mathcal{R}' und \mathcal{D} von \mathcal{S} ableitbar sind, folgt nach Lemma 4.5, daß \mathcal{R}'' von \mathcal{S} ableitbar ist. □

Somit besitzt \mathcal{R}'' bis auf die Minimalität alle Eigenschaften eines Reconciled Trees von \mathcal{G} über \mathcal{S} .

Aussage 12 Es gilt $|\mathcal{R}''| < |\mathcal{R}|$.

Beweis. Entsprechend den Ersetzungen \mathcal{R}' und \mathcal{R}'' errechnen wir die Kardinalität von \mathcal{R}'' . Es gilt

$$\begin{aligned}
|\mathcal{R}''| &= |\mathcal{R}| - |T_{\mathcal{R}}(X'')| + |T_S(B(X'')) \times p| - |T_{\mathcal{R}'}(\sqcup_S(X) \times \{p\})| + |\mathcal{D}| \\
&= |\mathcal{R}| - |T_{\mathcal{R}}(X'')| + |T_S(B(X''))| - |T_S(\sqcup_S(X))| + |\mathcal{D}| \\
&= |\mathcal{R}| - |T_{\mathcal{R}}(X'')| + |T_S(B(X''))| - |T_S(\sqcup_S(X))| + |\{\text{Ro}(\mathcal{D})\}| \\
&\quad + \sum_{i \in \{1, \dots, n\}} |T_{\mathcal{R}}(Y_i'')| \\
&= |\mathcal{R}| - c
\end{aligned}$$

mit $c := |\mathsf{T}_{\mathcal{R}}(\mathsf{X}'')| - |\mathsf{T}_{\mathcal{S}}(\mathsf{B}(\mathsf{X}''))| + |\mathsf{T}_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(\mathsf{X}))| - |\{\mathsf{Ro}(\mathcal{D})\}| - \sum_{i \in \{1, \dots, n\}} |\mathsf{T}_{\mathcal{R}}(\mathsf{Y}_i'')|$.
Wir zeigen $c \geq 1$. Es gilt

$$\begin{aligned}
& |\mathsf{T}_{\mathcal{R}}(\mathsf{X}'')| - \sum_{i \in \{1, \dots, n\}} |\mathsf{T}_{\mathcal{R}}(\mathsf{Y}_i'')| \\
&= |\{\mathsf{X}''\}| + \sum_{i \in \{1, \dots, n\}} (|\mathsf{T}_{\mathcal{R}}(\mathsf{C}_i'') \setminus \mathsf{T}_{\mathcal{R}}(\mathsf{Y}_i'')|) \\
&\geq 1 + \sum_{i \in \{1, \dots, n\}} (|\mathsf{B}(\mathsf{T}_{\mathcal{R}}(\mathsf{C}_i'') \setminus \mathsf{T}_{\mathcal{R}}(\mathsf{Y}_i''))|) \\
&\geq 1 + \sum_{i \in \{1, \dots, n\}} (|\mathsf{B}(\mathsf{T}_{\mathcal{R}}(\mathsf{C}_i'')) \setminus \mathsf{B}(\mathsf{T}_{\mathcal{R}}(\mathsf{Y}_i''))|) \\
&= 1 + \sum_{i \in \{1, \dots, n\}} (|\mathsf{B}(\mathsf{T}_{\mathcal{R}}(\mathsf{C}_i''))| - |\mathsf{B}(\mathsf{T}_{\mathcal{R}}(\mathsf{Y}_i''))|) \\
&= 1 + \sum_{i \in \{1, \dots, n\}} (|\mathsf{T}_{\mathcal{S}}(\mathsf{B}(\mathsf{C}_i''))| - |\mathsf{T}_{\mathcal{S}}(\mathsf{B}(\mathsf{Y}_i''))|) \tag{4.10}
\end{aligned}$$

$$= 1 + \sum_{i \in \{1, \dots, n\}} (|\mathsf{T}_{\mathcal{S}}(\mathsf{B}(\mathsf{X}''))| - |\mathsf{T}_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(\mathsf{X}))|) \tag{4.11}$$

$$\geq 1 + 2(|\mathsf{T}_{\mathcal{S}}(\mathsf{B}(\mathsf{X}''))| - |\mathsf{T}_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(\mathsf{X}))|) \tag{4.12}$$

$$\geq 5 \tag{4.13}$$

Zu (4.10): Sei $i \in \{1, \dots, n\}$. Da \mathcal{R} ein Reconciled Tree ist, sind $\mathsf{T}_{\mathcal{R}}(\mathsf{C}_i'')$ und $\mathsf{T}_{\mathcal{R}}(\mathsf{Y}_i'')$ von \mathcal{S} ableitbar. Daraus folgt, $\mathsf{B}(\mathsf{T}_{\mathcal{R}}(\mathsf{Y}_i'')) = \mathsf{T}_{\mathcal{S}}(\mathsf{B}(\mathsf{Y}_i''))$ und $\mathsf{B}(\mathsf{T}_{\mathcal{R}}(\mathsf{C}_i'')) = \mathsf{T}_{\mathcal{S}}(\mathsf{B}(\mathsf{C}_i''))$.

Zu (4.11): Nach Aussage 4 gilt $\mathsf{B}(\mathsf{C}_i'') = \mathsf{B}(\mathsf{X}'')$, und nach Definition von Y'' gilt $\mathsf{B}(\mathsf{Y}_i'') = \bigsqcup_{\mathcal{S}}(\mathsf{X})$, für alle $i \in \{1, \dots, n\}$.

Zu (4.12): Da \mathcal{R} ein n -Clustersystem ist, besitzt X'' mindestens zwei Kinder $\mathsf{C}_1'', \mathsf{C}_2''$, woraus $n \geq 2$ folgt.

Zu (4.13): Laut Annahme gilt $\bigsqcup_{\mathcal{S}}(\mathsf{X}) \neq \mathsf{B}(\mathsf{X}'')$, woraus $\bigsqcup_{\mathcal{S}}(\mathsf{X}) \subset \mathsf{B}(\mathsf{X}'')$ und damit $|\mathsf{T}_{\mathcal{S}}(\mathsf{B}(\mathsf{X}''))| - |\mathsf{T}_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(\mathsf{X}))| \geq 2$ folgt.

Also folgt

$$\begin{aligned}
c &= |\mathsf{T}_{\mathcal{R}}(\mathsf{X}'')| - \sum_{i \in \{1, \dots, n\}} (|\mathsf{T}_{\mathcal{R}}(\mathsf{Y}_i'')|) - |\mathsf{T}_{\mathcal{S}}(\mathsf{B}(\mathsf{X}''))| + |\mathsf{T}_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(\mathsf{X}))| - |\{\mathsf{Ro}(\mathcal{D})\}| \\
&\geq 5 - |\mathsf{T}_{\mathcal{S}}(\mathsf{B}(\mathsf{X}''))| + |\mathsf{T}_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(\mathsf{X}))| - 1 \\
&= 4 - (|\mathsf{T}_{\mathcal{S}}(\mathsf{B}(\mathsf{X}''))| - |\mathsf{T}_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(\mathsf{X}))|) \\
&\geq 2 \tag{4.14}
\end{aligned}$$

Zu (4.14): Wie (4.13).

Da $c \geq 1$, folgt somit $|\mathcal{R}''| < |\mathcal{R}|$. □

Also ist unter der Voraussetzung, daß \mathcal{R} ein Reconciled Tree ist, auch \mathcal{R}'' ein Reconciled Tree mit $|\mathcal{R}''| < |\mathcal{R}|$. Dies ist ein Widerspruch zu der vorausgesetzten Minimalität von \mathcal{R} . □

Verw. von: Lem. 3.4, 4.17, 9.1, 4.2, 4.5, 4.6, Kor. 4.1, Theo. 4.1, 4.2; Verw. in: Lem. 4.9, 4.13, 5.2, Theo 4.6, 4.8.

4.4 Die Topologie des Reconciled Tree

In diesem Abschnitt werden wir unter Verwendung der Operationen auf Reconciled Trees aus Abschnitt 4.1 und dem LCA-Theorem aus Abschnitt 4.3 die Topologie eines Reconciled Tree $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$ vollständig beschreiben. Dabei werden wir feststellen, daß sich die Topologie von \mathcal{R} nur durch die Topologie von \mathcal{G} und Subtopologien von \mathcal{S} beschreiben läßt. Damit läßt sich die Topologie des Reconciled Tree von \mathcal{R} unabhängig von diesem und eindeutig beschreiben. In Kapitel 5 werden wir mit dieser Kenntnis für \mathcal{G} und \mathcal{S} einen biologisch interpretierbaren und einfach zu berechnenden konstruktiven Reconciled Tree definieren, welcher dieselbe Topologie wie \mathcal{R} besitzt.

Wir beschreiben die Topologie von \mathcal{R} , indem wir diesen in Teilregionen partitionieren und deren Topologien zeigen. Durch die Einbettung des Genbaumes \mathcal{G} in den Reconciled Tree \mathcal{R} wird dieser in Regionen partitioniert, welche wir als Entwicklungsumgebungen von \mathcal{R} bezeichnen. Beispielsweise unterteilt in Abbildung 4.8 der in den Reconciled Tree \mathcal{R} eingebettete Genbaum \mathcal{G} diesen in die Entwicklungsumgebungen $E_{\mathcal{R}}(a), E_{\mathcal{R}}(b), E_{\mathcal{R}}(c), E_{\mathcal{R}}(1), E_{\mathcal{R}}(2), E_{\mathcal{R}}(3)$.

In Abschnitt 4.9 werden wir die Entwicklungsumgebungen in \mathcal{R} definieren und allgemeine Struktureigenschaften dieser Entwicklungsumgebungen zeigen. Eine Entwicklungsumgebung partitionieren wir in Abschnitt 4.4.3 in weitere Teilregionen, zu denen wir strukturelle Eigenschaften zeigen. Mit diesen gezeigten Eigenschaften werden wir abschließend in Abschnitt 4.4.3 die Topologie einer beliebigen Entwicklungsumgebung von \mathcal{R} zeigen.

4.4.1 Die Entwicklungsumgebungen von \mathcal{G}

Die globale Struktur von \mathcal{R} wird durch den Genbaum \mathcal{G} vorgegeben. Dieser steht mit \mathcal{R} durch eine minimale spezieskonforme Einbettung $\varphi \in \text{Emon}(\mathcal{G}, \mathcal{S})$ in Beziehung. Die Einbettung von \mathcal{G} in \mathcal{R} durch φ gibt die globale Struktur

von \mathcal{R} vor und unterteilt \mathcal{R} in Regionen unbekannter Topologie, welche wir als Entwicklungsumgebungen bezeichnen. Eine Entwicklungsumgebung von $X \in \mathcal{G}$ wird durch die Elemente in \mathcal{R} , welche in $T_{\mathcal{R}}(\varphi(X))$ und nicht in einem Mengensystem $T_{\mathcal{R}}(\text{Ch}(\varphi(Z)))$ mit $Z \in \text{Ch}(X)$ enthalten sind, gebildet.

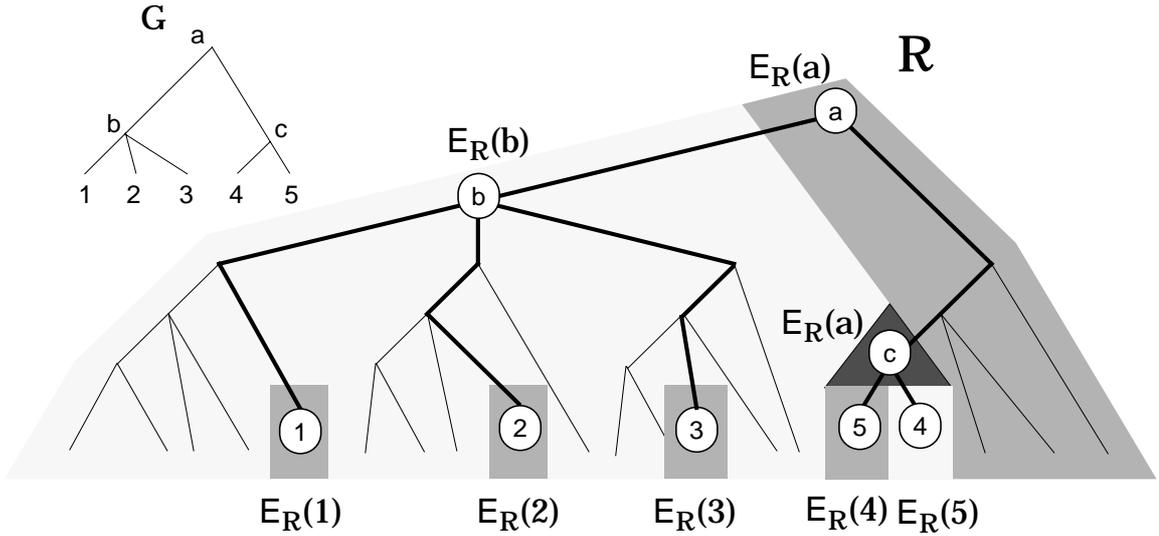


Abbildung 4.8: Die Entwicklungsumgebungen von \mathcal{R}

Definition 4.9 (Entwicklungsumgebung in \mathcal{R})

Sei $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$ ein Reconciled Tree von \mathcal{G} über \mathcal{S} , φ ein minimaler spezieskonformer Einbettungsmonomorphismus von \mathcal{G} in \mathcal{R} und $X \in \mathcal{G}$. Dann bezeichnen wir

$$E_{\mathcal{R}, \varphi}(X) := T_{\mathcal{R}}(\varphi(X)) \setminus \bigcup_{Z \in \text{Ch}(X)} T_{\mathcal{R}}(\varphi(Z))$$

als die Entwicklungsumgebung von X in \mathcal{R} . Wir verwenden die Bezeichnung $E_{\mathcal{R}}(X)$, wenn φ offensichtlich ist.

Lemma 4.7 (Partitionierung von \mathcal{R})

Es gilt $\mathcal{R} = \bigcup_{X \in \mathcal{G}} E_{\mathcal{R}}(X)$.

Beweis. Da φ ein Einbettungsmonomorphismus ist, gilt $E_{\mathcal{R}}(X) \cap E_{\mathcal{R}}(Y) = \emptyset$ für ungleiche $X, Y \in \mathcal{G}$. Somit ist die Vereinigung von $E_{\mathcal{R}}(X)$ und $E_{\mathcal{R}}(Y)$ disjunkt. Wir zeigen $\mathcal{R} \subseteq \bigcup_{X \in \mathcal{G}} E_{\mathcal{R}}(X)$. Dazu zeigen wir $\text{Ro}(\mathcal{R}) = \varphi(\text{Ro}(\mathcal{G}))$ durch Widerspruch. Nehmen wir $\text{Ro}(\mathcal{R}) \neq \varphi(\text{Ro}(\mathcal{G}))$ an, womit $\varphi(\text{Ro}(\mathcal{G})) \subset \text{Ro}(\mathcal{R})$ folgt. Aus $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$ und $\varphi(\text{Ro}(\mathcal{G})) \subset \text{Ro}(\mathcal{R})$ folgt $T_{\mathcal{R}}(\varphi(\text{Ro}(\mathcal{G}))) \in \text{RC}(\mathcal{G}, \mathcal{S})$. Damit gilt $|T_{\mathcal{R}}(\varphi(\text{Ro}(\mathcal{G})))| < |\mathcal{R}|$. Aber alle Reconciled Trees in $\text{RC}(\mathcal{G}, \mathcal{S})$ haben die gleiche Kardinalität, was ein Widerspruch ist.

Sei $Y'' \in \mathcal{R}$. Mit $\text{Ro}(\mathcal{R}) = \varphi(\text{Ro}(\mathcal{G}))$ folgt, daß ein $V \in \mathcal{G}$ mit $Y'' \subseteq \varphi(V)$ existiert. Da \mathcal{R} eine Baumordnung und φ injektiv ist, existiert genau ein minimales X mit $Y'' \subseteq \varphi(X)$. Somit ist $Y'' \in E_{\mathcal{R}}(X)$ und damit $\mathcal{R} \subseteq \bigcup_{X \in \mathcal{G}} E_{\mathcal{R}}(X)$. Die Aussage $\bigcup_{X \in \mathcal{G}} E_{\mathcal{R}}(X) \subseteq \mathcal{R}$ gilt offensichtlich. \square

Verw. von: ; Verw. in: Lem. 5.1, 5.2.

Abbildung 4.8 zeigt alle Entwicklungsumgebungen von \mathcal{R} . Im folgenden werden wir die Topologie einer beliebigen Entwicklungsumgebung $E_{\mathcal{R}}(X)$ mit $X \in \mathcal{G}$ beschreiben.

In dem folgenden Abschnitt 4.4.1.1 zeigen wir, daß zwischen einer bestimmten Region $E_{\mathcal{S}}(X)$ in \mathcal{S} und \mathcal{S} eine topologische Beziehung besteht. In Abschnitt 4.4.1.2 zeigen wir, welche Duplikationselemente $E_{\mathcal{R}}(X)$ enthalten kann, falls $E_{\mathcal{R}}(X)$ keine Duplikationen beinhaltet.

4.4.1.1 Der topologische Bezug von $E_{\mathcal{R}}(X)$ zu \mathcal{S}

Da \mathcal{R} von \mathcal{S} ableitbar ist, stehen die Topologien von $E_{\mathcal{R}}(X)$ und \mathcal{S} durch B in Bezug zueinander. Damit erhalten wir Aussagen über die Topologie von $E_{\mathcal{R}}(X)$ durch die Abbildungseigenschaften von B , bezogen auf $E_{\mathcal{R}}(X)$, also von $B_{E_{\mathcal{R}}(X)}$. Da $E_{\mathcal{R}}(X)$ offensichtlich unter \sqcup abgeschlossen ist, folgt aus dem folgenden Lemma, daß $B_{E_{\mathcal{R}}(X)}$ unter \sqcup abgeschlossen ist.

Lemma 4.8

Für jede unter \sqcup abgeschlossene Teilmenge $\mathcal{R}' \subseteq \mathcal{R}$ gilt $B_{\mathcal{R}'} \in \text{Ehom}(\mathcal{R}', \mathcal{S})$.

Beweis. Seien $V, W \in \mathcal{R}'$. Da \mathcal{R}' gegen \sqcup abgeschlossen ist, gilt $V \sqcup W \in \mathcal{R}'$, und es existiert in \mathcal{R}' genau ein maximales Element $\max(\mathcal{R}')$. Es gilt:

$$\begin{aligned} B_{\mathcal{R}'}(V \sqcup W) &= B_{\mathcal{R}'}(\min(K_{\mathcal{R}}(\max(\mathcal{R}'), V) \cap K_{\mathcal{R}}(\max(\mathcal{R}'), W))) \\ &= \min(B_{\mathcal{R}'}(K_{\mathcal{R}}(\max(\mathcal{R}'), V)) \cap B_{\mathcal{R}'}(K_{\mathcal{R}}(\max(\mathcal{R}'), W))) \\ &= \min(K_{\mathcal{S}}(B_{\mathcal{R}'}(\max(\mathcal{R}')), B_{\mathcal{R}'}(V)) \sqcup K_{\mathcal{S}}(B_{\mathcal{R}'}(\max(\mathcal{R}')), B_{\mathcal{R}'}(W))) \\ &= B_{\mathcal{R}'}(W) \sqcup B_{\mathcal{R}'}(V) \end{aligned}$$

Also gilt $B_{\mathcal{R}'} \in \text{Ehom}(\mathcal{R}', B(\mathcal{R}'))$. \square

Verw. von: ; Verw. in: Theo. 4.5, 4.8.

Der Wertebereich von $B_{E_{\mathcal{R}}(X)}$ beschreibt die Host-Spezies aller Gene aus $E_{\mathcal{R}}(X)$. Damit existiert zu X (neben $E_{\mathcal{R}}(X)$) auch eine Entwicklungsumgebung in \mathcal{S} , welche die Host-Spezies der Gene beschreibt, die durch die Entwicklung von X zu all seinen Kindern $\text{Ch}(X)$ entstanden sind. Analog zu der Entwicklungsumgebung in \mathcal{R} definieren wir die Entwicklungsumgebung in \mathcal{S} .

Definition 4.10 (Entwicklungsumgebung in \mathcal{S})

Sei M eine Menge und $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$, wobei \mathcal{G} ein Genbaum und \mathcal{S} ein Speziesbaum ist. Dann bezeichnen wir

$$E_{\mathcal{S}}(X) = T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \setminus \bigcup_{Z \in \text{Ch}(X)} T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z))$$

als die Entwicklungsumgebung von X in \mathcal{S} .

Die Gene in $E_{\mathcal{R}}(X)$ benötigen für deren Entwicklung mindestens alle Spezies aus $E_{\mathcal{S}}(X)$, wie das folgende Lemma zeigt.

Lemma 4.9

Sei $X \in \mathcal{G}$, dann gilt $B(E_{\mathcal{R}}(X)) \supseteq E_{\mathcal{S}}(X)$.

Beweis. Es gilt

$$\begin{aligned} B(E_{\mathcal{R}}(X)) &= B(T_{\mathcal{R}}(\varphi(X)) \setminus \bigcup_{Z \in \text{Ch}(X)} T_{\mathcal{R}}(\varphi(Z))) \\ &\supseteq B(T_{\mathcal{R}}(\varphi(X))) \setminus \bigcup_{Z \in \text{Ch}(X)} B(T_{\mathcal{R}}(\varphi(Z))) \quad \text{Lem. 3.3} \\ &= T_{\mathcal{S}}(B(\varphi(X))) \setminus \bigcup_{Z \in \text{Ch}(X)} T_{\mathcal{S}}(B(\varphi(Z))) \quad \mathcal{R} \text{ ist aus } \mathcal{S} \text{ ableitbar.} \\ &= T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \setminus \bigcup_{Z \in \text{Ch}(X)} T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z)) \quad \text{LCA-Theo. 4.4} \\ &= E_{\mathcal{S}}(X) \end{aligned}$$

□

Verw. von: Theo. 4.4, Lem. 3.3; *Verw. in:* Lem. 4.14, Theo. 4.5, Kor. 4.3.

4.4.1.2 Duplikationselemente in $E_{\mathcal{R}}(X)$

Wir zeigen, welche Duplikationselemente $E_{\mathcal{R}}(X)$ beinhaltet, wenn $|B(E_{\mathcal{R}}(X))| = |E_{\mathcal{R}}(X)|$ gilt. $E_{\mathcal{R}}(X)$ ist nur ohne Duplikationen von $B(E_{\mathcal{R}}(X))$ ableitbar, wenn $|B(E_{\mathcal{R}}(X))| = |E_{\mathcal{R}}(X)|$ gilt. Dies bedeutet jedoch nicht, daß $E_{\mathcal{R}}(X)$ kein Duplikationselement oder keine Kopie einer Duplikation beinhalten kann. Denn $E_{\mathcal{R}}(X)$ kann die Kopie $\varphi(X)$ des Duplikationselementes $\text{Pa}(\varphi(X))$ und Duplikationselemente $\text{Pa}(\varphi(Z))$ für $Z \in \text{Ch}(X)$ beinhalten, obwohl $B(E_{\mathcal{R}}(X)) = E_{\mathcal{R}}(X)$ gilt. Beispielsweise kann $|B(E_{\mathcal{R}}(X))| = |E_{\mathcal{R}}(X)|$ gelten und $\varphi(X)$ Duplikationselement und Kopie in $E_{\mathcal{R}}(X)$ sein, wenn $E_{\mathcal{R}}(X) = \{\varphi(X)\}$ gilt. $\varphi(X)$ ist eine Kopie von $\text{Pa}(\varphi(X))$, wenn ein weiteres Kind $X_i'' \in \text{Ch}(\text{Pa}(\varphi(X)))$ mit $B(\varphi(X)) \cap B(X_i'') \neq \emptyset$

existiert. $\varphi(X)$ ist ein Duplikationselement, wenn $B(\varphi(Z_i)) \cap B(\varphi(Z_j)) \neq \emptyset$ für ungleiche $Z_i, Z_j \in \text{Ch}(X)$ gilt.¹

Jedoch kann eine Duplikation nicht "vollständig" in $E_{\mathcal{R}}(X)$ stattfinden.

Lemma 4.10

Sei $X \in \mathcal{G}$ und $|B(E_{\mathcal{R}}(X))| = |E_{\mathcal{R}}(X)|$, genau dann existieren keine $Y'', Z'' \in E_{\mathcal{R}}(X)$, mit $Y'' \in \text{Dup}_{\mathcal{R}}$ und $Z'' \in \text{Ch}(Y'')$.

Beweis. Wir zeigen: Existieren $Y'', Z'' \in E_{\mathcal{R}}(X)$ mit $Y'' \in \text{Dup}_{\mathcal{R}}$ und $Z'' \in \text{Ch}(Y'')$, so folgt $|B(E_{\mathcal{R}}(X))| \neq |E_{\mathcal{R}}(X)|$. Aus $Y'' \in \text{Dup}_{\mathcal{R}}$ und $Z'' \in \text{Ch}(Y'')$ folgt $B(Y'') = B(Z'')$. Damit gilt

$$|E_{\mathcal{R}}(X)| - 1 = |E_{\mathcal{R}}(X) \setminus \{Y''\}| \geq |B(E_{\mathcal{R}}(X) \setminus \{Y''\})| = |B(E_{\mathcal{R}}(X))|.$$

Also gilt $|E_{\mathcal{R}}(X)| - 1 \geq |B(E_{\mathcal{R}}(X))|$, woraus $|B(E_{\mathcal{R}}(X))| \neq |E_{\mathcal{R}}(X)|$ folgt.

Wir zeigen: Existieren keine $Y'', Z'' \in E_{\mathcal{R}}(X)$ mit $Y'' \in \text{Dup}_{\mathcal{R}}$ und $Z'' \in \text{Ch}(Y'')$, so folgt $|B(E_{\mathcal{R}}(X))| = |E_{\mathcal{R}}(X)|$. Da keine $Y'', Z'' \in E_{\mathcal{R}}(X)$ mit $Y'' \in \text{Dup}_{\mathcal{R}}$ und $Z'' \in \text{Ch}(Y'')$ existieren, folgt $B(A) \cap B(B) = \emptyset$ für alle $A, B \in E_{\mathcal{R}}(X)$. Somit folgt $|B(E_{\mathcal{R}}(X))| = |E_{\mathcal{R}}(X)|$. \square

Wenn $|E_{\mathcal{R}}(X)| = |B(E_{\mathcal{R}}(X))|$ gilt, kann somit nur das Maximum von $E_{\mathcal{R}}(X)$ eine Kopie sein, und es können nur die minimalen Elemente von $E_{\mathcal{R}}(X)$ Duplikationselemente sein. Beinhaltet $E_{\mathcal{R}}(X)$ ein Duplikationselement und eine seiner Kopien, so bezeichnen wir dies als ein *vollständiges Duplikationsereignis* in $E_{\mathcal{R}}(X)$.²

4.4.2 Regionen einer Entwicklungsumgebung

Wir werden in diesem Kapitel eine Entwicklungsumgebung $E_{\mathcal{R}}(X)$ mit $X \in \mathcal{G}$ in weitere Regionen, die lokalen Entwicklungsumgebungen und die unbestimmte Entwicklungsumgebung von $E_{\mathcal{R}}(X)$, partitionieren.

$E_{\mathcal{R}}(X)$ beschreibt in \mathcal{R} die Entwicklung des Gens X zu allen seinen Kindern, wenn diese existent sind. Damit beschreibt $E_{\mathcal{R}}(X)$ auch die *lokale* Entwicklung von X zu einem Kind $Z \in \text{Ch}(X)$, welche wir als lokale Entwicklungsumgebung von Z bezeichnen.

¹Es gilt $\varphi(Z_j), \varphi(Z_i) \notin E_{\mathcal{R}}(X)$.

²Da wir diese Eigenschaft nur im sprachlichen Teil und nicht für unseren Formalismus verwenden, verzichten wir hier auf eine formale Definition.

Definition 4.11 (Lokale Entwicklungsumgebung von $E_{\mathcal{R}}(X)$)

Sei $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$ ein Reconciled Tree von \mathcal{G} über \mathcal{S} , φ ein minimaler spezieskonformer Einbettungsmonomorphismus von \mathcal{G} in \mathcal{R} und $Z \in \mathcal{G}$. Dann bezeichnen wir

$$e_{\mathcal{R}, \varphi}(Z) := \{Y'' \in \mathcal{R} \mid \exists C'' \in \text{Ch}(\varphi(\text{Pa}(Z))) : \\ \varphi(Z) \subset C'' \Rightarrow Y'' \in T_{\mathcal{R}}(C'') \setminus T_{\mathcal{R}}(\varphi(Z))\}$$

als die lokale Entwicklungsumgebung von Z in $E_{\mathcal{R}, \varphi}(X)$. Wir verwenden die Bezeichnung $e_{\mathcal{R}}(Z)$, wenn φ offensichtlich ist.

Abbildung 4.9 zeigt, als Beispiel, alle lokalen Entwicklungsumgebungen von $E_{\mathcal{R}}(b)$. Hierbei sind $b'' = \varphi(b)$ sowie $1'' := \varphi(1)$, $2'' := \varphi(2)$ und $3'' := \varphi(3)$.

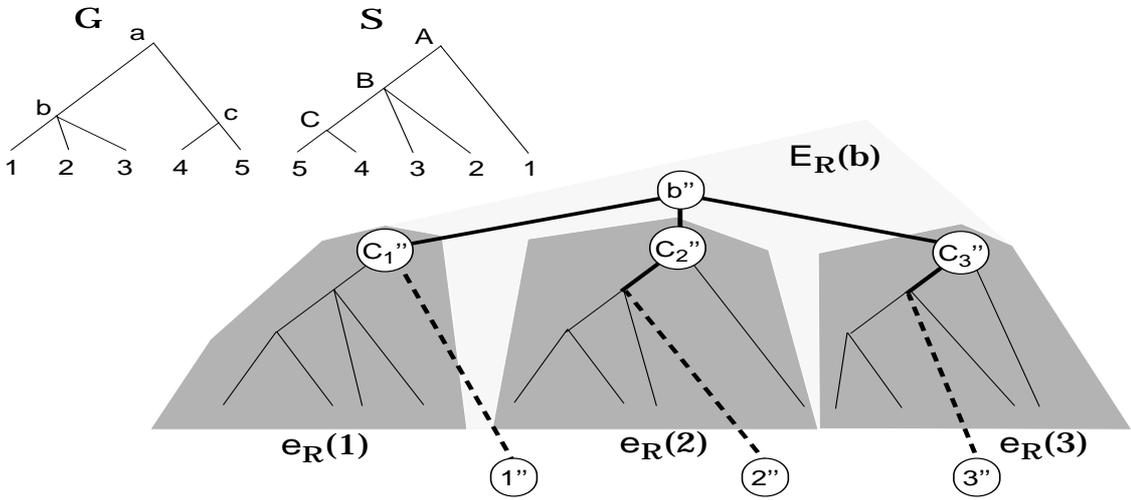


Abbildung 4.9: Die lokalen Entwicklungsumgebungen von $E_{\mathcal{R}}(b)$

Möglicherweise existieren Elemente in $E_{\mathcal{R}}(X) \setminus \{\varphi(X)\}$, welche sich in keiner lokalen Entwicklungsumgebung $e_{\mathcal{R}}(Z)$ mit $Z \in \text{Ch}(X)$ befinden. Diese Elemente sind wie bei der lokalen Entwicklungsumgebung durch die Entwicklung von X zu dessen Kindern entstanden. Durch die Multifurkation sind diese jedoch den Kindern nicht mehr eindeutig zuordenbar und bilden die im folgenden definierte unbestimmte Entwicklungsumgebung von X in \mathcal{R} .

Definition 4.12 (Unbestimmte Entwicklungsumgebung von $E_{\mathcal{R}}(X)$)

Sei $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$ ein Reconciled Tree von \mathcal{G} über \mathcal{S} , φ ein minimaler spezieskonformer Einbettungsmonomorphismus von \mathcal{G} in \mathcal{R} und $Z \in \mathcal{G}$. Dann bezeichnen wir

$$\tilde{E}_{\mathcal{R}}(X) := E_{\mathcal{R}}(X) \setminus \{\varphi(X)\} \cup \bigcup_{Z \in \text{Ch}(X)} e_{\mathcal{S}}(Z)$$

als die unbestimmte Entwicklungsumgebung von $E_{\mathcal{R}}(X)$.

Nach Abbildung 4.9 gilt, $\tilde{E}_{\mathcal{R}}(\mathbf{b}) = \emptyset$. Wir werden später in Theorem 4.7 zeigen, daß dies für jede Entwicklungsumgebung eines Reconciled Tree gilt, wenn \mathbf{b} ein Duplikationselement ist.

Lemma 4.11 (Partition von $E_{\mathcal{R}}(X)$)

Sei $X \in \mathcal{G}$, dann gilt $E_{\mathcal{R}}(X) = \{\varphi(X)\} \dot{\cup} \tilde{E}_{\mathcal{R}}(X) \dot{\cup} \bigcup_{Z \in \text{Ch}(X)} e_{\mathcal{R}}(Z)$.

Beweis. Offensichtlich gilt $E_{\mathcal{R}}(X) = \{\varphi(X)\} \cup \tilde{E}_{\mathcal{R}}(X) \cup \bigcup_{Z \in \text{Ch}(X)} e_{\mathcal{R}}(Z)$. Seien $Z_i, Z_j \in \text{Ch}(X)$ mit $Z_i \neq Z_j$. Da φ einbettungsmonomorph ist, folgt $\varphi(Z_i) \cap \varphi(Z_j) = \emptyset$ und $\varphi(Z_i) \subset \varphi(X)$. Damit ist die Vereinigung von $\varphi(Z_i)$, $\varphi(Z_j)$ und $\varphi(X)$ disjunkt.

Lokale Entwicklungsumgebungen und die unbestimmte Entwicklungsumgebung von $E_{\mathcal{R}}(X)$ sowie $\{\varphi(X)\}$ sind nach Definition dieser Umgebungen paarweise disjunkt. \square

Verw. von: ; Verw. in: Theo. 4.7.

Eine lokale Entwicklungsumgebung $e_{\mathcal{R}}(Z)$ von $E_{\mathcal{R}}(X)$ mit $Z \in \text{Ch}(X)$ enthält jedoch kein Duplikationselement, wenn $|e_{\mathcal{R}}(Z)| = |B(e_{\mathcal{R}}(Z))|$ gilt.

Lemma 4.12

Sei $X \in \mathcal{G}, Z \in \text{Ch}(X)$ und $|B(e_{\mathcal{R}}(Z))| = |e_{\mathcal{R}}(Z)|$, genau dann existiert kein $Y'' \in e_{\mathcal{R}}(Z)$ mit $Y'' \in \text{Dup}_{\mathcal{R}}$.

Beweis. Sei $Y'' \in e_{\mathcal{R}}(Z)$ ein Duplikationselement, womit $B(Y_i'') = B(Y'')$ für jedes Kind $Y_i'' \in \text{Ch}(Y'')$ gilt. Im Gegensatz zu einem Element in $E_{\mathcal{R}}(X)$ besitzt Y'' mindestens ein Kind Y_i'' , welches auch in $e_{\mathcal{R}}(Z)$ enthalten ist. Damit ist der Beweis analog zu dem Beweis von Lemma 4.10. \square

Verw. von: ; Verw. in: Lem. 4.17.

Der Bezug unter der Abbildung $B_{E_{\mathcal{R}}(X)}$ von $E_{\mathcal{R}}(X)$ zu $E_{\mathcal{S}}(X)$ überträgt sich auch auf die lokalen Entwicklungsumgebungen und die unbestimmte Entwicklungsumgebung von $E_{\mathcal{R}}(X)$, wie wir im folgenden Abschnitt zeigen werden.

4.4.2.1 Der Bezug der Regionen von $E_{\mathcal{R}}(X)$ zu $E_{\mathcal{S}}(X)$

Auch eine lokale Entwicklungsumgebung $e_{\mathcal{R}}(Z)$ mit $Z \in \text{Ch}(X)$ und die unbestimmte Entwicklungsumgebung $\tilde{E}_{\mathcal{R}}(X)$ besitzen, wie $E_{\mathcal{R}}(X)$ zu $E_{\mathcal{S}}(X)$, durch B einen topologischen Bezug zu einer bestimmten Region in $E_{\mathcal{S}}(X)$. Wir unterteilen daher auch $E_{\mathcal{S}}(Z)$ in lokale Regionen und in eine unbestimmte Region, in welchen $B(e_{\mathcal{R}}(Z))$ und $B(\tilde{E}_{\mathcal{R}}(X))$ jeweils enthalten sind. Dazu definieren wir zunächst analog zu der lokalen und unbestimmten Entwicklungsumgebung in $E_{\mathcal{R}}(X)$ eine lokale und unbestimmte Entwicklungsumgebung in $E_{\mathcal{S}}(X)$.

Definition 4.13 (Lokale Entwicklungsumgebung von $E_S(X)$)

Seien M eine Menge, $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$ wobei \mathcal{G} ein Genbaum und \mathcal{S} ein Speziesbaum ist, sowie $Z \in \mathcal{G}$. Dann bezeichnen wir

$$e_S(Z) := \{Y' \in \mathcal{S} \mid \exists C' \in \text{Ch}(\bigsqcup_S(\text{Pa}(Z))) : \\ \bigsqcup_S(Z) \subset C' \Rightarrow Y' \in T_S(C') \setminus T_S(\bigsqcup_S(Z))\}$$

als die lokale Entwicklungsumgebung von Z in $E_S(X)$.

Definition 4.14 (Unbestimmte Entwicklungsumgebung von $E_S(X)$)

Sei $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$ ein Reconciled Tree von \mathcal{G} über \mathcal{S} , φ ein minimaler spezieskonformer Einbettungsmonomorphismus von \mathcal{G} in \mathcal{R} und $Z \in \mathcal{G}$. Dann bezeichnen wir

$$\tilde{E}_S(X) := E_S(X) \setminus (\{\bigsqcup_S(X)\} \cup \bigcup_{Z \in \text{Ch}(X)} e_S(Z))$$

als die unbestimmte Entwicklungsumgebung von $E_S(X)$.

Wir haben in Lemma 4.9 gezeigt, daß die Spezies $E_S(X)$ in den Host-Spezies $B(E_{\mathcal{R}}(X))$ der Gene $E_{\mathcal{R}}(X)$ enthalten sind. Wir zeigen nun, daß sich diese Eigenschaft auf die lokalen Entwicklungsumgebungen in $e_S(Z)$ und $e_{\mathcal{R}}(X)$ sowie $\tilde{E}_{\mathcal{R}}(X)$ und $\tilde{E}_S(X)$ überträgt.

Lemma 4.13

Sei $X \in \mathcal{G}$ und $Z \in \text{Ch}(X)$, dann gilt $B(e_{\mathcal{R}}(Z)) \supseteq e_S(Z)$.

Beweis. Sei C'' das Maximum von $e_{\mathcal{R}}(Z)$ und C' das Maximum von $e_S(Z)$. Es gilt

$$\begin{aligned} B(e_{\mathcal{R}}(Z)) &= B(T_{\mathcal{R}}(C'') \setminus T_{\mathcal{R}}(\varphi(Z))) \\ &\supseteq B(T_{\mathcal{R}}(C'') \setminus B(T_{\mathcal{R}}(\varphi(Z)))) \quad \text{Lemma 3.3} \\ &= T_S(B(C'')) \setminus T_S(B(\varphi(Z))) \quad \mathcal{R} \text{ ist aus } \mathcal{S} \text{ ableitbar.} \\ &= T_S(B(C'')) \setminus T_S(\bigsqcup_S(Z)) \quad \text{LCA-Theorem 4.4} \\ &\supseteq T_S(C') \setminus T_S(\bigsqcup_S(Z)) \quad \text{folgt aus } C' \subseteq B(C'') \\ &= e_S(Z) \end{aligned}$$

Wir zeigen abschließend $C' \subseteq B(C'')$: Es gilt $C' \in \text{Ch}(\bigsqcup_S(X))$ und $C'' \in \text{Ch}(\varphi(X))$.

Wir zeigen $B(C'') \in \text{Ch}(\bigsqcup_S(X))$ oder $B(C'') = \bigsqcup_S(X)$. Aus $C'' \in \text{Ch}(\varphi(X))$ folgt $B(C'') = B(\varphi(X))$, wenn $\varphi(X)$ ein Duplikationselement ist, und $B(C'') \in$

$\text{Ch}(B(\varphi(X)))$ sonst. Nach dem LCA-Theorem 4.4 ist $B(\varphi(X)) = \sqcup_{\mathcal{S}}(X)$. Somit folgt $B(C'') \in \text{Ch}(\sqcup_{\mathcal{S}}(X))$ oder $B(C'') = \sqcup_{\mathcal{S}}(X)$.

Gilt $B(C'') = \sqcup_{\mathcal{S}}(X)$: Mit $C' \in \text{Ch}(\sqcup_{\mathcal{S}}(X))$ folgt $C' \in B(C'')$, womit $C' \subset B(C'')$ gilt. Gilt $B(C'') \in \text{Ch}(\sqcup_{\mathcal{S}}(X))$: Es gilt $\sqcup_{\mathcal{S}}(Z) \subseteq C'$ und $\varphi(Z) \subseteq C''$. Aus $\varphi(Z) \subseteq C''$ folgt $B(\varphi(Z)) \subseteq B(C'')$. Mit dem LCA-Theorem 4.4 folgt weiter $\sqcup_{\mathcal{S}}(Z) \subseteq B(C'')$. Also gilt $\sqcup_{\mathcal{S}}(Z) \subseteq C'$ und $\sqcup_{\mathcal{S}}(Z) \subseteq B(C'')$, womit C' und $B(C'')$ durch \subseteq in \mathcal{S} vergleichbar sind. Da $C' \in \text{Ch}(\sqcup_{\mathcal{S}}(X))$ und $B(C'') \in \text{Ch}(\sqcup_{\mathcal{S}}(X))$ gilt, sind $B(C'')$ und C' nur vergleichbar, wenn sie gleich sind. Somit gilt $C' = B(C'')$. \square

Verw. von: Theo. 4.4, Lem. 3.3; Verw. in: Lem. 4.14, Kor. 4.3.

Lemma 4.14

Sei $X \in \mathcal{G}$, dann gilt $B(\tilde{E}_{\mathcal{R}}(X)) \supseteq \tilde{E}_{\mathcal{S}}(X)$.

Beweis.

$$\begin{aligned}
B(\tilde{E}_{\mathcal{R}}(X)) &= B(E_{\mathcal{R}}(X) \setminus (\{\varphi(X)\} \cup \bigcup_{Z \in \text{Ch}(X)} e_{\mathcal{R}}(Z))) \\
&\supseteq B(E_{\mathcal{R}}(X)) \setminus (B(\{\varphi(X)\}) \cup \bigcup_{Z \in \text{Ch}(X)} B(e_{\mathcal{R}}(Z))) && \text{Lem. 3.3} \\
&\supseteq E_{\mathcal{S}}(X) \setminus (B(\{\varphi(X)\}) \cup \bigcup_{Z \in \text{Ch}(X)} B(e_{\mathcal{R}}(Z))) && \text{Lem. 4.9} \\
&\supseteq E_{\mathcal{S}}(X) \setminus (B(\{\varphi(X)\}) \cup \bigcup_{Z \in \text{Ch}(X)} e_{\mathcal{S}}(Z)) && \text{Lem. 4.13} \\
&= E_{\mathcal{S}}(X) \setminus (\sqcup_{\mathcal{S}}(X) \cup \bigcup_{Z \in \text{Ch}(X)} e_{\mathcal{S}}(Z)) && \text{Theo. 4.4} \\
&= \tilde{E}_{\mathcal{S}}(X)
\end{aligned}$$

\square

Verw. von: Lem. 3.3, 4.9, 4.13, Theo. 4.4; Verw. in: Kor. 4.3.

Im Gegensatz zu $E_{\mathcal{R}}(X)$ wird $E_{\mathcal{S}}(X)$ im allgemeinen nicht durch seine lokalen Entwicklungsumgebungen $e_{\mathcal{S}}(Z)$ mit $Z \in \text{Ch}(X)$ und die unbestimmte Entwicklungsumgebung $\tilde{E}_{\mathcal{S}}(X)$ partitioniert.

Lemma 4.15

Sei $X \in \mathcal{G}$, dann gilt $E_{\mathcal{S}}(X) \subseteq \sqcup_{\mathcal{S}}(X) \cup \tilde{E}_{\mathcal{S}}(X) \cup \bigcup_{Z \in \text{Ch}(X)} e_{\mathcal{S}}(Z)$.

Beweis. Wenn ungleiche $Z_i, Z_j \in \text{Ch}(X)$ mit $e_{\mathcal{S}}(Z_i) \cap e_{\mathcal{S}}(Z_j) \neq \emptyset$ existieren, so gilt $E_{\mathcal{S}}(X) \subset \sqcup_{\mathcal{S}}(X) \cup \tilde{E}_{\mathcal{S}}(X) \cup \bigcup_{Z \in \text{Ch}(X)} e_{\mathcal{S}}(Z)$. Sonst gilt $E_{\mathcal{S}}(X) = \sqcup_{\mathcal{S}}(X) \cup \tilde{E}_{\mathcal{S}}(X) \cup \bigcup_{Z \in \text{Ch}(X)} e_{\mathcal{S}}(Z)$. \square

Verw. von: ; Verw. in: .

4.4.3 Die Topologie einer Entwicklungsumgebung $E_{\mathcal{R}}(X)$

Nach der Untergliederung von \mathcal{R} in Regionen und den Aussagen über lokale Eigenschaften dieser Regionen besitzen wir alle Voraussetzungen, um die Topologie einer Entwicklungsumgebung $E_{\mathcal{R}}(X)$ mit $X \in \mathcal{G}$ zu zeigen.

In Abschnitt 4.4.3.1 zeigen wir, daß $E_{\mathcal{R}}(X)$ und $E_{\mathcal{S}}(X)$ dieselbe Topologie besitzen, wenn das Prädikat $P_{\text{NoDup}}(X)$ (welches wir in dem nächsten Abschnitt definieren) gilt. Hieraus leiten wir dann ab, daß $e_{\mathcal{R}}(Z)$ und $e_{\mathcal{S}}(Z)$ für jedes $Z \in \text{Ch}(X)$ sowie $\tilde{E}_{\mathcal{R}}(X)$ und $\tilde{E}_{\mathcal{S}}(X)$ jeweils dieselbe Topologie besitzen.

In Abschnitt 4.4.3.2 untersuchen wir die Auswirkungen des Prädikates $P_{\text{NoDup}}(X)$ auf die Duplikationselemente in $E_{\mathcal{R}}(X)$. Wir zeigen durch das Duplikationstheorem, daß genau dann, wenn $P_{\text{NoDup}}(X)$ nicht gilt, die Entwicklungsumgebung $E_{\mathcal{R}}(X)$ Duplikationselemente enthält. Enthält $E_{\mathcal{R}}(X)$ Duplikationselemente, so ist $\varphi(X)$ ein Duplikationselement. Im besonderen reduziert sich das Duplikationstheorem – bei der Anwendung auf bifurkierende Phylogenien – auf das von Page verwendete, aber nicht gezeigte Kriterium, um Duplikationselemente zu bestimmen. Wir zeigen dies in Korollar 4.4.

In Abschnitt 4.4.3.3 zeigen wir dann die Topologie von $E_{\mathcal{R}}(X)$, wenn $P_{\text{NoDup}}(X)$ nicht gilt. Nach dem Duplikationstheorem ist $\varphi(X)$ ein Duplikationselement in $E_{\mathcal{R}}(X)$. Wir zeigen, daß jedes Duplikat von $E_{\mathcal{R}}(X)$ die Wurzel genau einer lokalen Entwicklungsumgebung $e_{\mathcal{R}}(Z)$ mit $Z \in \text{Ch}(X)$ ist. Damit beinhaltet $E_{\mathcal{R}}(X)$ keine unbestimmte Entwicklungsumgebung $\tilde{E}_{\mathcal{R}}(X)$. Die Topologie einer Entwicklungsumgebung $e_{\mathcal{R}}(Z)$ entspricht der Topologie von $E_{\mathcal{S}}(X) \setminus e_{\mathcal{S}}(Z)$, welches wir durch $B_{e_{\mathcal{R}}(Z)} \in \text{Oiso}(e_{\mathcal{R}}(Z), E_{\mathcal{S}}(X) \setminus e_{\mathcal{S}}(Z))$ mit $Z \in \text{Ch}(X)$ zeigen.

Damit besitzt $E_{\mathcal{R}}(X)$ in Abhängigkeit von der Gültigkeit des Prädikates $P_{\text{NoDup}}(X)$ eine von zwei möglichen Topologietypen, welche sich aus Subtopologien von \mathcal{S} zusammensetzen.

4.4.3.1 Die Topologie von $E_{\mathcal{R}}(X)$, wenn $P_{\text{NoDup}}(X)$ gilt

In diesem Abschnitt werden wir zunächst $B_{E_{\mathcal{R}}(X)} \in \text{Oiso}(E_{\mathcal{R}}(X), E_{\mathcal{S}}(X))$ unter dem wahren Prädikat $P_{\text{NoDup}}(X)$, welches wir noch definieren werden, zeigen. Dann werden wir hieraus $B_{\tilde{E}_{\mathcal{R}}(X)} \in \text{Oiso}(\tilde{E}_{\mathcal{R}}(X), \tilde{E}_{\mathcal{S}}(X))$ und $B_{e_{\mathcal{R}}(Z)} \in \text{Oiso}(e_{\mathcal{R}}(Z), e_{\mathcal{S}}(Z))$ für ein beliebiges $Z \in \text{Ch}(X)$ ableiten.

Bevor wir die Aussage $B_{E_{\mathcal{R}}(X)} \in \text{Oiso}(E_{\mathcal{R}}(X), E_S(X))$ formal unter dem wahren Prädikat $P_{\text{NoDup}}(X)$ zeigen, skizzieren wir zunächst die Idee des Beweises und leiten dabei die Definition des Prädikates $P_{\text{NoDup}}(X)$ ab.

Grundlage des Beweises ist die Konstruktion eines Reconciled Tree $\mathcal{R}' \in \text{RC}(\mathcal{G}, \mathcal{S})$, in welchen \mathcal{G} mit dem minimalen Einbettungsmonomorphismus φ' spezieskonform einbettbar ist.³ \mathcal{R}' besitzt dabei die beiden folgenden Eigenschaften: $|E_{\mathcal{R}'}(X)| = |E_S(X)|$ und $|E_{\mathcal{R}}(X)| = |E_{\mathcal{R}'}(X)|$. Hieraus folgt die wesentliche Aussage

$$|E_{\mathcal{R}}(X)| = |E_S(X)|.$$

Wir wollen $B_{E_{\mathcal{R}}(X)} \in \text{Oiso}(E_{\mathcal{R}}(X), E_S(X))$ zeigen. Lemma 4.8 besagt, daß $B_{E_{\mathcal{R}}(X)} \in \text{Ehom}(E_{\mathcal{R}}(X), \mathcal{S})$ und damit $B_{E_{\mathcal{R}}(X)} \in \text{Ohom}(E_{\mathcal{R}}(X), \mathcal{S})$ gilt. Also gilt auch $B_{E_{\mathcal{R}}(X)} \in \text{Oepi}(E_{\mathcal{R}}(X), B(E_{\mathcal{R}}(X)))$. Die Aussage $|E_{\mathcal{R}}(X)| = |E_S(X)|$ verwenden wir nun einerseits, um $B(E_{\mathcal{R}}(X)) = E_S(X)$ zu zeigen, womit $B_{E_{\mathcal{R}}(X)} \in \text{Oepi}(E_{\mathcal{R}}(X), E_S(X))$ folgt. Andererseits folgt mit $|E_{\mathcal{R}}(X)| = |E_S(X)|$ nun die zu zeigende Aussage $B_{E_{\mathcal{R}}(X)} \in \text{Eiso}(E_{\mathcal{R}}(X), E_S(X))$.

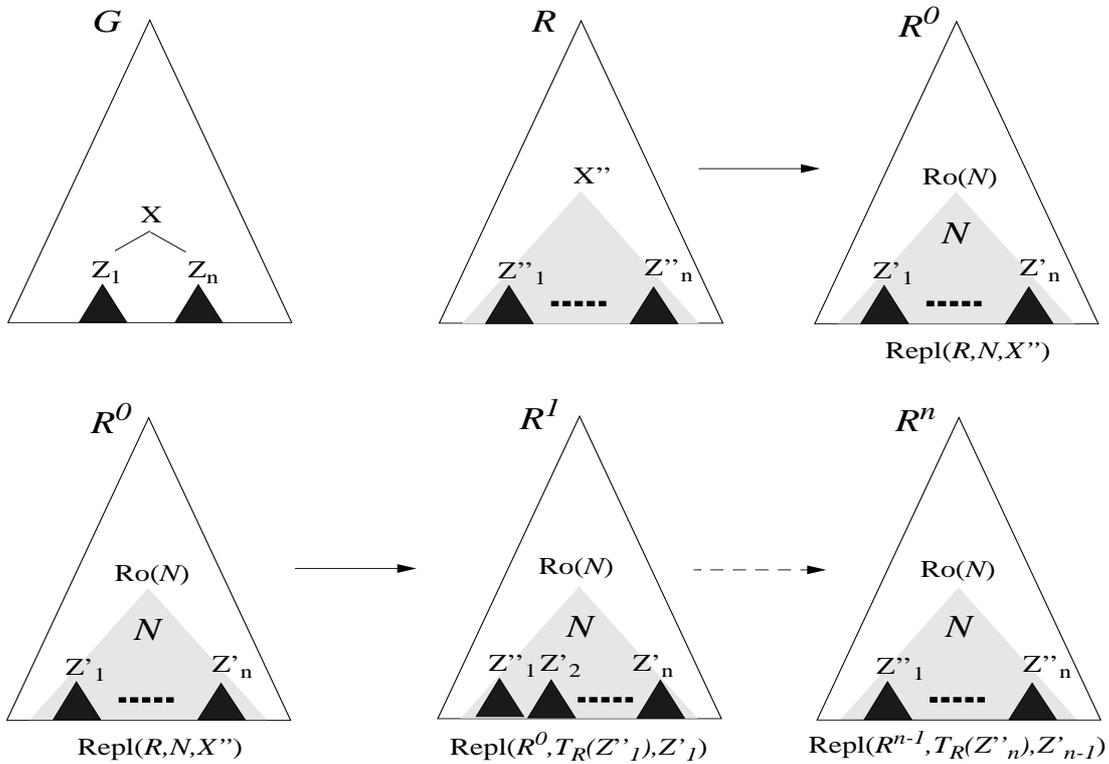


Abbildung 4.10: Die Konstruktion von \mathcal{R}^n

Konstruieren wir nun \mathcal{R}' mit Hilfe der Operationen Repl und \times . Abbildung 4.10 veranschaulicht die Konstruktion von $\mathcal{R}' = \mathcal{R}^n$. Das Element $X \in \mathcal{G}$ wird durch

³ $E_{\mathcal{R}'}$ ist damit die vereinfachte Form von $E_{\mathcal{R}', \varphi'}$.

φ auf $X'' := \varphi(X)$ und dessen Kinder $Z_1, \dots, Z_n \in \text{Ch}(X)$ auf die Elemente $Z_1'' := \varphi(Z_1), \dots, Z_n'' := \varphi(Z_n)$ in \mathcal{R} abgebildet. Wir ersetzen zuerst den Teilbaum $T_{\mathcal{R}}(X'')$ durch die duplikationsfreie Ableitung $\mathcal{N} := T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \times q$ von $T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X))$ und erhalten die Ableitung \mathcal{R}^0 von \mathcal{S} . Damit \mathcal{N} durch Repl ersetzt werden darf, wählen wir $q \in \mathbb{N}$ tdisjunkt zu $\text{Ro}(\mathcal{R})$.

Um die spezieskonforme Einbettbarkeit von \mathcal{G} in \mathcal{R}^0 sicherzustellen, ersetzen wir schrittweise in \mathcal{R}^0 jeden Teilbaum gewurzelt an einem $Z_i' := \bigsqcup_{\mathcal{S}}(Z_i) \times \{q\}$ mit $i \in \{1, \dots, n\}$ durch den Teilbaum $T_{\mathcal{R}}(\varphi(Z_i))$. Im ersten Schritt ersetzen wir in \mathcal{R}^0 den Teilbaum $T_{\mathcal{R}^0}(Z_i')$ durch den Teilbaum $T_{\mathcal{R}}(Z_i')$ und erhalten \mathcal{R}^1 . In Schritt $i \in \{2, \dots, n\}$ ersetzen wir in \mathcal{R}^{i-1} den Teilbaum $T_{\mathcal{R}^{i-1}}(Z_i')$ durch den Teilbaum $T_{\mathcal{R}}(Z_i'')$ und erhalten \mathcal{R}^i . Um die spezieskonforme Einbettbarkeit von \mathcal{G} sicherzustellen, darf nicht durch die Ersetzung eines Teilbaumes $T_{\mathcal{R}^{i-1}}(Z_i')$ ein schon ersetzter Teilbaum $T_{\mathcal{R}}(\varphi(Z_j))$ mit $j \in \{1, \dots, i-1\}$ ersetzt werden. Dies sichern wir durch Erzwingung des Prädikates $P_{\text{NoDup}}(X)$.

Definition 4.15 ($P_{\text{NoDup}}(X)$)

Sei $X \in \mathcal{G}$. Das Prädikat $P_{\text{NoDup}}(X)$ ist definitionsgemäß genau dann wahr, wenn für alle $Z_i, Z_j \in \text{Ch}(X)$ mit $i \neq j$ gilt:

$$\bigsqcup_{\mathcal{S}}(Z_i) \subset \bigsqcup_{\mathcal{S}}(X) \text{ und } \bigsqcup_{\mathcal{S}}(Z_i) \sqcup \bigsqcup_{\mathcal{S}}(Z_j) = \bigsqcup_{\mathcal{S}}(X)$$

Damit ist $\mathcal{R}' = \mathcal{R}^n$ ein Reconciled Tree. Für \mathcal{R}' gilt: Die Topologien von $\mathcal{R}' \setminus E_{\mathcal{R}'}(X)$ und $\mathcal{R} \setminus E_{\mathcal{R}}(X)$ sind gleich und $|E_{\mathcal{R}'}(X)| = |E_{\mathcal{S}}(X)|$. Aus der Gleichheit der Topologien von $\mathcal{R}' \setminus E_{\mathcal{R}'}(X)$ und $\mathcal{R} \setminus E_{\mathcal{R}}(X)$ folgt $|\mathcal{R}' \setminus E_{\mathcal{R}'}(X)| = |\mathcal{R} \setminus E_{\mathcal{R}}(X)|$. Da $\mathcal{R}, \mathcal{R}' \in \text{RC}(\mathcal{G}, \mathcal{S})$ gilt, folgt $|\mathcal{R}| = |\mathcal{R}'|$. Aus $|\mathcal{R}' \setminus E_{\mathcal{R}'}(X)| = |\mathcal{R}'|$ und $|\mathcal{R}| = |\mathcal{R}'|$ folgt dann $|E_{\mathcal{R}}(X)| = |E_{\mathcal{R}'}(X)|$. Somit erfüllt \mathcal{R}' die geforderten Eigenschaften $|E_{\mathcal{R}'}(X)| = |E_{\mathcal{S}}(X)|$ und $|E_{\mathcal{R}}(X)| = |E_{\mathcal{R}'}(X)|$.

Lemma 4.16

Sei $X \in \mathcal{G}$, dann gilt $P_{\text{NoDup}}(X)$ genau dann, wenn für alle $Z_i, Z_j \in \text{Ch}(X)$ mit $i \neq j$ gilt,

$$B(\varphi(Z_i)) \subset B(\varphi(X)) \text{ und } B(\varphi(Z_i)) \sqcup B(\varphi(Z_j)) = B(\varphi(X)).$$

Das folgende Theorem zeigt formal, daß unter der Bedingung P_{NoDup} (4.15) die Entwicklungsumgebung $E_{\mathcal{R}}(X)$ eine duplikationsfreie Ableitung aus \mathcal{S} ist und damit keine vollständigen Duplikationsereignisse enthält.

Theorem 4.5

Sei $X \in \mathcal{G}$ und $P_{\text{NoDup}}(X)$ wahr, dann gilt:

$$B_{E_{\mathcal{R}}(X)} \in \text{Oiso}(E_{\mathcal{R}}(X), E_{\mathcal{S}}(X)).$$

Beweis. Wir zeigen zuerst, daß $|E_{\mathcal{R}}(X)| = |E_{\mathcal{S}}(X)|$ gilt. Dazu definieren wir für X und dessen Kinder Z_1, \dots, Z_n mit $n \in \mathbb{N}$, basierend auf \mathcal{R} , das Multimensystem \mathcal{R}^n . Für ein zu $\text{Ro}(\mathcal{R})$ tdisjunktes $p \in \mathbb{N}$ definieren wir

$$\mathcal{R}^0 := \text{Repl}(\mathcal{R}, T_{\mathcal{S}}(B(\varphi(X))) \times p, \varphi(X)) \text{ und}$$

$$\mathcal{R}^i := \text{Repl}(\mathcal{R}^{i-1}, T_{\mathcal{R}}(\varphi(Z_i)), B(\varphi(Z_i)) \times \{p\}) \text{ für alle } i \in \{1, \dots, n\}.$$

Da $P_{\text{NoDup}}(X)$ gilt, folgt $\mathcal{R}^n \in \text{RC}(\mathcal{G}, \mathcal{S})$ aus Lemma 9.3.

Die Kardinalität von \mathcal{R}^n berechnet sich anhand der Kardinalität von \mathcal{R} wie folgt:

$$\begin{aligned} |\mathcal{R}^n| &= |\mathcal{R}| - |T_{\mathcal{R}}(\varphi(X))| + |T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \times p| \\ &\quad + \sum_{i \in \{1, \dots, n\}} (-|T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z_i)) \times p| + |T_{\mathcal{R}}(\varphi(Z_i))|) \\ &= |\mathcal{R}| - |T_{\mathcal{R}}(\varphi(X))| + |T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X))| \\ &\quad + \sum_{Z \in \text{Ch}(X)} (-|T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z))| + |T_{\mathcal{R}}(\varphi(Z))|) \\ &= |\mathcal{R}| - (|T_{\mathcal{R}}(\varphi(X))| - \sum_{Z \in \text{Ch}(X)} |T_{\mathcal{R}}(\varphi(Z))|) \\ &\quad + |T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X))| - \sum_{Z \in \text{Ch}(X)} |T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z))| \\ &= |\mathcal{R}| - (|T_{\mathcal{R}}(\varphi(X))| - | \bigcup_{Z \in \text{Ch}(X)} T_{\mathcal{R}}(\varphi(Z)) |) \\ &\quad + |T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X))| - | \bigcup_{Z \in \text{Ch}(X)} T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z)) | \tag{4.15} \\ &= |\mathcal{R}| - (|T_{\mathcal{R}}(\varphi(X)) \setminus \bigcup_{Z \in \text{Ch}(X)} T_{\mathcal{R}}(\varphi(Z)) |) \\ &\quad + |T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \setminus \bigcup_{Z \in \text{Ch}(X)} T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z)) | \tag{4.16} \\ &= |\mathcal{R}| - |E_{\mathcal{R}}(X)| + |E_{\mathcal{S}}(X)|. \end{aligned}$$

Zu (4.15): Es gilt $| \bigcup_{Z \in \text{Ch}(X)} T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z)) | = \sum_{Z \in \text{Ch}(X)} |T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z))|$, da $\bigsqcup_{\mathcal{S}}(Z_i) \cap \bigsqcup_{\mathcal{S}}(Z_j) = \emptyset$ für alle ungleichen $Z_i, Z_j \in \text{Ch}(X)$ gilt. Angenommen, es existieren

ungleiche $Z_i, Z_j \in \text{Ch}(X)$ mit $\sqcup_S(Z_i) \cap \sqcup_S(Z_j) \neq \emptyset$, so folgt in der Baumordnung von \mathcal{S} , daß o.B.d.A. $\sqcup_S(Z_i) \subset \sqcup_S(Z_j)$ gilt. Damit ist $\sqcup_S(Z_i) \sqcup \sqcup_S(Z_j) = \sqcup_S(Z_j)$ und nach Voraussetzung gilt $\sqcup_S(Z_j) \subset \sqcup_S(X)$. Nach Voraussetzung gilt jedoch auch $\sqcup_S(Z_i) \sqcup \sqcup_S(Z_j) = \sqcup_S(X)$, was ein Widerspruch ist.

Es gilt $|\bigcup_{Z \in \text{Ch}(X)} T_{\mathcal{R}}(\varphi(Z))| = \sum_{Z \in \text{Ch}(X)} |T_{\mathcal{R}}(\varphi(Z))|$, da $\varphi(Z_i) \cap \varphi(Z_j) = \emptyset$ für alle ungleichen $Z_i, Z_j \in \text{Ch}(X)$ gilt. Dies folgt aus der Einbettungsmonomorphie von φ .

Zu (4.15): Die Umrechnung in $T_{\mathcal{S}}(\sqcup_S(X))$ gilt, da $\sqcup_S(Z) \subseteq \sqcup_S(X)$ für alle $Z \in \text{Ch}(X)$ gilt, woraus $T_{\mathcal{S}}(\sqcup_S(Z)) \subseteq T_{\mathcal{S}}(\sqcup_S(X))$ für jedes $Z \in \text{Ch}(X)$ folgt. Die Umrechnung in $T_{\mathcal{R}}(\varphi(X))$ gilt, da $\varphi(Z) \subset \varphi(X)$ für alle $Z \in \text{Ch}(X)$, aufgrund der Einbettungsmonomorphie von φ , gilt. Damit folgt $T_{\mathcal{R}}(\varphi(Z)) \subset T_{\mathcal{R}}(\varphi(X))$ für jedes $Z \in \text{Ch}(X)$.

Da $\mathcal{R}, \mathcal{R}^n \in \text{RC}(\mathcal{G}, \mathcal{S})$ gilt, folgt $|\mathcal{R}| = |\mathcal{R}^n|$ (aufgrund der Minimalität beider m -Clustersysteme). Somit folgt

$$|E_{\mathcal{R}}(X)| = |E_{\mathcal{S}}(X)|.$$

Da $E_{\mathcal{R}}(X)$ gegen \sqcup abgeschlossen ist, folgt $B_{E_{\mathcal{R}}(X)} \in \text{Ehom}(E_{\mathcal{R}}(X), \mathcal{S})$ aus Lemma 4.8 und weiter $B_{E_{\mathcal{R}}(X)} \in \text{Ohom}(E_{\mathcal{R}}(X), \mathcal{S})$. Also gilt auch

$$B_{E_{\mathcal{R}}(X)} \in \text{Oepi}(E_{\mathcal{R}}(X), B_{E_{\mathcal{R}}(X)}(E_{\mathcal{R}}(X))).$$

Wir zeigen, daß $B(E_{\mathcal{R}}(X)) = E_{\mathcal{S}}(X)$ gilt. Nach Lemma 4.9 gilt, $B(E_{\mathcal{R}}(X)) \supseteq E_{\mathcal{S}}(X)$. Daraus folgt $|B(E_{\mathcal{R}}(X))| \geq |E_{\mathcal{S}}(X)|$. Es gilt $|E_{\mathcal{R}}(X)| \geq |B(E_{\mathcal{R}}(X))|$, und wir haben gerade $|E_{\mathcal{S}}(X)| = |E_{\mathcal{R}}(X)|$ gezeigt. Also gilt

$$|E_{\mathcal{R}}(X)| \geq |B(E_{\mathcal{R}}(X))| \geq |E_{\mathcal{S}}(X)| = |E_{\mathcal{R}}(X)|,$$

woraus $|B(E_{\mathcal{R}}(X))| = |E_{\mathcal{S}}(X)|$ folgt. Mit $B(E_{\mathcal{R}}(X)) \supseteq E_{\mathcal{S}}(X)$ folgt hieraus $B(E_{\mathcal{R}}(X)) = E_{\mathcal{S}}(X)$.

Somit ist $B_{E_{\mathcal{R}}(X)} \in \text{Oepi}(E_{\mathcal{R}}(X), E_{\mathcal{S}}(X))$, und da $|E_{\mathcal{R}}(X)| = |E_{\mathcal{S}}(X)|$ gilt, folgt $B_{E_{\mathcal{R}}(X)} \in \text{Oiso}(E_{\mathcal{R}}(X), E_{\mathcal{S}}(X))$. \square

Verw. von: Lem. 9.3, 4.8, 4.9; Verw. in: Lem. 4.17, Kor. 4.3, Theo. 4.6.

Korollar 4.3

Sei $X \in \mathcal{G}$, $P_{\text{NoDup}}(X)$ wahr und $Z \in \text{Ch}(X)$. Dann gilt

$$B_{e_{\mathcal{R}}(Z)} \in \text{Oiso}(e_{\mathcal{R}}(Z), e_{\mathcal{S}}(Z)) \text{ und } B_{\tilde{E}_{\mathcal{R}}(X)} \in \text{Oiso}(\tilde{E}_{\mathcal{R}}(X), \tilde{E}_{\mathcal{S}}(X)).$$

Beweis. Nach Theorem 4.5 gilt $B_{E_{\mathcal{R}}(X)} \in \text{Oiso}(E_{\mathcal{R}}(X), E_S(X))$, woraus $B_{e_{\mathcal{R}}(Z)} \in \text{Oiso}(e_{\mathcal{R}}(Z), B(e_{\mathcal{R}}(Z)))$ und $B_{\tilde{E}_{\mathcal{R}}(X)} \in \text{Oiso}(\tilde{E}_{\mathcal{R}}(X), B(\tilde{E}_{\mathcal{R}}(X)))$ folgt. Wenn $e_S(Z) = B(e_{\mathcal{R}}(Z))$ und $\tilde{E}_S(X) = B(\tilde{E}_{\mathcal{R}}(X))$ gilt, folgt die zu zeigende Aussage.

Wir beweisen also noch $e_S(Z) = B(e_{\mathcal{R}}(Z))$ und $\tilde{E}_S(X) = B(\tilde{E}_{\mathcal{R}}(X))$. Dazu zeigen wir zunächst

$$|B(\tilde{E}_{\mathcal{R}}(X))| + \sum_{Y \in \text{Ch}(X)} |B(e_{\mathcal{R}}(Y))| = |\tilde{E}_S(X)| + \sum_{Y \in \text{Ch}(X)} |e_S(Y)|.$$

$$\begin{aligned} |E_{\mathcal{R}}(X)| &= |\bigsqcup_S(X)| + |\tilde{E}_{\mathcal{R}}(X)| + \sum_{Y \in \text{Ch}(X)} |e_{\mathcal{R}}(Y)| \\ &\geq 1 + |B(\tilde{E}_{\mathcal{R}}(X))| + \sum_{Y \in \text{Ch}(X)} |B(e_{\mathcal{R}}(Y))| \\ &\geq 1 + |\tilde{E}_S(X)| + \sum_{Y \in \text{Ch}(X)} |e_{\mathcal{R}}(Y)| && \text{Lem. 4.13, 4.14} \\ &\geq |E_S(X)| \\ &= |E_{\mathcal{R}}(X)| \end{aligned}$$

Daraus folgt $|B(\tilde{E}_{\mathcal{R}}(X))| + \sum_{Y \in \text{Ch}(X)} |B(e_{\mathcal{R}}(Y))| = |\tilde{E}_S(X)| + \sum_{Y \in \text{Ch}(X)} |e_S(Y)|$.

Es gilt $e_S(Y) \subseteq B(e_{\mathcal{R}}(Y))$ für jedes $Y \in \text{Ch}(X)$ nach Lem. 4.13 und $\tilde{E}_S(X) \subseteq B(\tilde{E}_{\mathcal{R}}(X))$ nach Lem. 4.14. Hieraus folgt $|e_S(Y)| \leq |B(e_{\mathcal{R}}(Y))|$ für jedes $Y \in \text{Ch}(X)$ und $|\tilde{E}_S(X)| \leq |B(\tilde{E}_{\mathcal{R}}(X))|$. Mit $|B(\tilde{E}_{\mathcal{R}}(X))| + \sum_{Y \in \text{Ch}(X)} |B(e_{\mathcal{R}}(Y))| = |\tilde{E}_S(X)| + \sum_{Y \in \text{Ch}(X)} |e_S(Y)|$ folgt hieraus $|e_S(Y)| = |B(e_{\mathcal{R}}(Y))|$ für jedes $Y \in \text{Ch}(X)$ und $|\tilde{E}_S(X)| = |B(\tilde{E}_{\mathcal{R}}(X))|$.

Somit gilt im besonderen $e_S(Z) = B(e_{\mathcal{R}}(Z))$ und $\tilde{E}_S(X) = B(\tilde{E}_{\mathcal{R}}(X))$. \square

Verw. von: Lem. 4.13, 4.14; Verw. in: .

4.4.3.2 Das Duplikations-Theorem

In diesem Abschnitt werden wir zeigen, daß $E_{\mathcal{R}}(X)$ genau dann Duplikationselemente besitzt, wenn $P_{\text{NoDup}}(X)$ nicht gilt. Bis jetzt können wir aus Theorem 4.5 und Lemma 4.10 nur folgern, daß $E_{\mathcal{R}}(X)$ kein vollständiges Duplikationsereignis beinhaltet, wenn $P_{\text{NoDup}}(X)$ gilt. Wir werden nun zeigen, daß $\varphi(X)$ ein Duplikationselement ist, wenn $E_{\mathcal{R}}(X)$ Duplikationselemente enthält (also P_{NoDup} nicht gilt). Damit können wir Aussagen über Duplikationselemente in einem Reconciled Tree formulieren, ohne diesen zu kennen.

Theorem 4.6 (Duplikations-Theorem)

Sei $X \in \mathcal{G}$. Genau dann, wenn $P_{\text{NoDup}}(X)$ nicht gilt, besitzt $E_{\mathcal{R}}(X)$ Duplikationselemente. Besitzt $E_{\mathcal{R}}(X)$ Duplikationselemente, so ist $\varphi(X)$ ein Duplikationselement.

Beweis.

Genau dann, wenn $P_{\text{NoDup}}(X)$ nicht gilt, existieren ungleiche $Z_i, Z_j \in \text{Ch}(X)$, für die $\sqcup_S(Z_i) = \sqcup_S(X)$ oder $\sqcup_S(Z_i) \sqcup \sqcup_S(Z_j) \subset \sqcup_S(X)$ gilt.

1. Wir zeigen: Wenn $\sqcup_S(Z_i) = \sqcup_S(X)$ oder $\sqcup_S(Z_i) \sqcup \sqcup_S(Z_j) \subset \sqcup_S(X)$ gilt, ist $\varphi(X)$ ein Duplikationselement.

- (a) Wir zeigen: Gilt $\sqcup_S(Z_i) = \sqcup_S(X)$, dann ist $\varphi(X)$ ein Duplikationselement. Aus $\sqcup_S(Z_i) = \sqcup_S(X)$ folgt mit dem LCA-Theorem 4.4, daß $B(\varphi(X)) = B(\varphi(Z_i))$ gilt. Damit gilt $B(Y'') = B(\varphi(X))$, für jedes $Y'' \in K_{\mathcal{R}}(\varphi(X), \varphi(Z_i))$. Im besonderen gilt $B(C_i'') = B(\varphi(X))$ für das Kind $C_i'' \in \text{Ch}(\varphi(X)) \cap K_{\mathcal{R}}(\varphi(X), \varphi(Z_i))$.

Da \mathcal{R} ein n -Clustersystem ist, existiert ein zweites Kind $C_k'' \in \text{Ch}(\varphi(X))$, mit $C_k'' \neq C_i''$. Aus $C_k'' \in \text{Ch}(\varphi(X))$ folgt $C_k'' \subset \varphi(X)$, und mit Lemma 3.1 folgt weiter $B(C_k'') \subseteq B(\varphi(X))$.

Es gilt damit $B(\varphi(X)) = B(C_i'')$ und $B(C_k'') \subseteq B(\varphi(X))$ für die ungleichen Kinder C_i'' und C_k'' von $\text{Ch}(\varphi(X))$, woraus $B(C_k) \cap B(C_i) \neq \emptyset$ folgt. Somit ist $\varphi(X)$ ein Duplikationselement.

- (b) Wir zeigen: Gilt $\sqcup_S(Z_i) \sqcup \sqcup_S(Z_j) \subset \sqcup_S(X)$, dann ist $\varphi(X)$ ein Duplikationselement.

Wir zeigen zuerst, daß ungleiche Kinder $C_i'', C_j'' \in \text{Ch}(\varphi(X))$ mit $B(C_i''), B(C_j'') \in K_S(\sqcup_S(X), \sqcup_S(Z_i) \sqcup \sqcup_S(Z_j))$ existieren. Mit dem LCA-Theorem 4.4 gilt:

$$\begin{aligned}
& B(K_{\mathcal{R}}(\varphi(X), \varphi(Z_i)) \cap B(K_{\mathcal{R}}(\varphi(X), \varphi(Z_j)))) \\
&= K_S(B(\varphi(X)), B(\varphi(Z_i))) \cap K_S(B(\varphi(X)), B(\varphi(Z_j))) \\
&= K_S\left(\sqcup_S(X), \sqcup_S(Z_i)\right) \cap K_S\left(\sqcup_S(X), \sqcup_S(Z_j)\right) \\
&= K_S\left(\sqcup_S(X), \sqcup_S(Z_i) \sqcup \sqcup_S(Z_j)\right).
\end{aligned}$$

Also existieren $Y_i'' \in K_{\mathcal{R}}(\varphi(X), \varphi(Z_i))$ und $Y_j'' \in K_{\mathcal{R}}(\varphi(X), \varphi(Z_j))$ mit $B(Y_i'') = B(Y_j'') = \sqcup_S(Z_i) \sqcup \sqcup_S(Z_j)$. Aus der Voraussetzung $\sqcup_S(Z_i) \sqcup \sqcup_S(Z_j) \subset \sqcup_S(X)$ folgt $B(Y_i''), B(Y_j'') \subset \sqcup_S(X)$. Nach dem LCA-Theorem 4.4 gilt $\sqcup_S(X) = B(\varphi(X))$, woraus $B(Y_i''), B(Y_j'') \subset B(\varphi(X))$

folgt. Aus $B(Y_i''), B(Y_j'') \subset B(\varphi(X))$ folgt $Y_i'' \neq \varphi(X)$ und $Y_j'' \neq \varphi(X)$. Damit folgt aus $Y_i'' \in K_{\mathcal{R}}(\varphi(X), \varphi(Z_i))$ und $Y_j'' \in K_{\mathcal{R}}(\varphi(X), \varphi(Z_j))$, daß $Y_i'', Y_j'' \subset \varphi(X)$ gilt. Somit existieren Kinder $C_i'', C_j'' \in \text{Ch}(\varphi(X))$ mit $C_i'' \in K_{\mathcal{R}}(\varphi(X), Y_i'')$ und $C_j'' \in K_{\mathcal{R}}(\varphi(X), Y_j'')$.

Wir zeigen $B(C_i'') \cap B(C_j'') \neq \emptyset$. Es gilt

$$B(C_i''), B(C_j'') \in K_S(\bigsqcup_S(X), \bigsqcup_S(Z_i) \sqcup \bigsqcup_S(Z_j)).$$

Dies folgt, da $B(C_i'') \in B(K_{\mathcal{R}}(\varphi(X), Y_i''))$ ist und

$$\begin{aligned} B(K_{\mathcal{R}}(\varphi(X), Y_i'')) &= K_S(B(\varphi(X)), B(Y_i'')) \\ &= K_S\left(\bigsqcup_S(X), \bigsqcup_S(Z_i) \sqcup \bigsqcup_S(Z_j)\right) \end{aligned}$$

gilt. Entsprechend folgert $B(C_j'') \in K_S(\bigsqcup_S(X), \bigsqcup_S(Z_i) \sqcup \bigsqcup_S(Z_j))$. Aus $B(C_i''), B(C_j'') \in K_S(\bigsqcup_S(X), \bigsqcup_S(Z_i) \sqcup \bigsqcup_S(Z_j))$ folgt $B(C_i'') \cap B(C_j'') \neq \emptyset$.

Da $C_i'' \in K_{\mathcal{R}}(\varphi(X), Y_i'')$, $C_j'' \in K_{\mathcal{R}}(\varphi(X), Y_j'')$ und $C_i'', C_j'' \in \text{Ch}(\varphi(X))$ gilt, folgt $C_i'' \neq C_j''$. Mit $B(C_i'') \cap B(C_j'') \neq \emptyset$ ist $\varphi(X)$ somit ein Duplikationselement.

2. Wir zeigen: Existieren keine ungleichen $Z_i, Z_j \in \text{Ch}(X)$ mit

$$\bigsqcup_S(Z_i) = \bigsqcup_S(X) \text{ oder } \bigsqcup_S(Z_i) \sqcup \bigsqcup_S(Z_j) \subset \bigsqcup_S(X),$$

so besitzt $E_{\mathcal{R}}(X)$ kein Duplikationselement.

Wir zeigen dies durch Widerspruch. Nehmen wir an, es existiert das Duplikationselement $D \in E_{\mathcal{R}}(X)$. Da D ein Duplikationselement ist, existieren Kinder $V, W \in \text{Ch}(D)$ mit $B(V) \cap B(W) \neq \emptyset$. Daraus folgt $B(V) = B(W) = B(D)$ in dem m -Clustersystem \mathcal{R} , womit $B(U) = B(D)$ für alle Kinder $U \in \text{Ch}(D)$ gilt.

Nach Theorem 4.5 gilt $B_{E_{\mathcal{R}}(X)} \in \text{Iso}(E_{\mathcal{R}}(X), E_S(X))$, womit B auf $E_{\mathcal{R}}(X)$ injektiv ist. Demzufolge existiert in $E_{\mathcal{R}}(X)$ kein Kind von D . Also ist zwar D , aber keines der Kinder von D $\text{Ch}(D)$ in $E_{\mathcal{R}}(X)$ enthalten. Damit existieren ungleiche $Z_i, Z_j \in \text{Ch}(X)$ mit $\varphi(Z_i), \varphi(Z_j) \in \text{Ch}(D)$. Für diese Kinder gilt $B(\varphi(Z_i)) = B(\varphi(Z_j)) = B(D)$, woraus $\bigsqcup_S(Z_i) = \bigsqcup_S(Z_j)$ mit dem LCA-Theorem 4.4 folgt. Damit gilt $\bigsqcup_S(Z_i) \sqcup \bigsqcup_S(Z_j) = \bigsqcup_S(Z_j)$. Da $\bigsqcup_S(Z_j) \subset \bigsqcup_S(X)$ gilt, folgt entweder $\bigsqcup_S(Z_i) = \bigsqcup_S(X)$ oder $\bigsqcup_S(Z_i) \sqcup \bigsqcup_S(Z_j) \subset \bigsqcup_S(X)$ im Widerspruch zu der Voraussetzung.

□

Verw. von: Theo. 4.4, 4.5, Lem. 3.1; Verw. in: Kor. 4.4, 4.5, Theo. 4.7, Lem. 4.17, 5.1.

Die Duplikationsbedingung des Duplikations-Theorems reduziert sich, wenn X nur zwei Kinder besitzt.

Korollar 4.4

Seien \mathcal{R} binär, $X \in \mathcal{G}$ und $\{Z_1, Z_2\} = \text{Ch}(X)$. $E_{\mathcal{R}}(X)$ besitzt genau dann ein Duplikationselement, wenn $\sqcup_{\mathcal{S}}(Z_1) = \sqcup_{\mathcal{S}}(X)$ oder $\sqcup_{\mathcal{S}}(Z_2) = \sqcup_{\mathcal{S}}(X)$ gilt. Besitzt $E_{\mathcal{R}}(X)$ Duplikationselemente, so ist $\varphi(X)$ ein Duplikationselement.

Beweis. Das Duplikations-Theorem 4.6 besagt, daß $E_{\mathcal{R}}(X)$ genau dann eine Duplikation besitzt, wenn $\sqcup_{\mathcal{S}}(Z_1) = \sqcup_{\mathcal{S}}(X)$, $\sqcup_{\mathcal{S}}(Z_2) = \sqcup_{\mathcal{S}}(X)$ oder $\sqcup_{\mathcal{S}}(Z_1) \sqcup \sqcup_{\mathcal{S}}(Z_2) \subset X$ gilt. Besitzt $E_{\mathcal{R}}(X)$ Duplikationselemente, so ist $\varphi(X)$ ein Duplikationselement.

Da $\{Z_1, Z_2\} = \text{Ch}(X)$, folgt $Z_1 \dot{\cup} Z_2 = X$ und damit gilt $\sqcup_{\mathcal{S}}(X) = \sqcup_{\mathcal{S}}(Z_1) \sqcup \sqcup_{\mathcal{S}}(Z_2)$. Somit existiert der Fall $\sqcup_{\mathcal{S}}(Z_1) \sqcup \sqcup_{\mathcal{S}}(Z_2) \subset \sqcup_{\mathcal{S}}(X)$ nicht, und die Aussage ist gezeigt. \square

Verw. von: Theo. 4.6; Verw. in: .

Page hat informal Korollar 4.4 verwendet, um den Reconciled Tree zu konstruieren. Er hat damit die Elemente in \mathcal{G} bestimmt, welche eingebettet in dem zu konstruierenden Reconciled Tree, Duplikationselemente sind. Mit diesen Elementen hat er dann nach einem rekursiven Verfahren den Reconciled Tree konstruiert. Da er nur binäre Bäume verwendet hat, obwohl seine Definition multifurkierende Bäume zuläßt, ist das von Page verwandte Kriterium ausreichend gewesen.

4.4.3.3 Die Topologie von $E_{\mathcal{R}}(X)$, wenn $\neg P_{\text{NoDup}}(X)$ gilt

Wir beschreiben die Topologie einer Entwicklungsumgebung $E_{\mathcal{R}}(X)$ $X \in \mathcal{G}$, wenn $P_{\text{NoDup}}(X)$ nicht gilt. Mit der Topologie von $E_{\mathcal{R}}(X)$ werden wir anschließend das Duplikations-Theorem 4.6 durch einen Korollar verschärfen. Wir zeigen, daß $\varphi(X)$ das einzige Duplikationselement in $\varphi(X)$ ist, wenn $E_{\mathcal{R}}(X)$ ein Duplikationselement beinhaltet.

Da $P_{\text{NoDup}}(X)$ nicht gilt, ist $\varphi(X)$ nach dem Duplikations-Theorem 4.6 ein Duplikationselement. Nach dem folgenden Theorem ist $E_{\mathcal{R}}(X)$ dann nur in seine lokalen Entwicklungsumgebungen und $\{\varphi(X)\}$ partitionierbar. Da $\varphi(X)$ die Wurzel von $E_{\mathcal{R}}(X)$ ist, sind dessen Duplikate Wurzeln von jeweils genau einer lokalen Entwicklungsumgebung $e_{\mathcal{R}}(Z)$ mit $Z \in \text{Ch}(X)$.

Kennen wir somit die Topologien der lokalen Entwicklungsumgebungen von $E_{\mathcal{R}}(X)$, so kennen wir auch die Topologie von $E_{\mathcal{R}}(X)$.

Theorem 4.7

Sei $X \in \mathcal{G}$. Wenn $E_{\mathcal{R}}(X)$ ein Duplikationselement beinhaltet, dann gilt

$$E_{\mathcal{R}}(X) = \{\varphi(X)\} \dot{\cup} \bigcup_{Z \in \text{Ch}(X)} e_{\mathcal{R}}(Z).$$

Beweis. Zur Vereinfachung definieren wir $\mathcal{P} := \{\varphi(X)\} \cup \bigcup_{Z \in \text{Ch}(X)} e_{\mathcal{R}}(Z)$. Aus Lemma 4.11 folgt $E_{\mathcal{R}}(X) = \mathcal{P} \cup \tilde{E}_{\mathcal{R}}(X)$, womit $\tilde{E}_{\mathcal{R}}(X) = \emptyset$ zu zeigen bleibt.

$E_{\mathcal{R}}(X)$ beinhaltet ein Duplikationselement. Nach dem Duplikations-Theorem 4.6 ist damit $\varphi(X)$ ein Duplikationselement. Nehmen wir an, es existiert ein $Y'' \in \tilde{E}_{\mathcal{R}}(X)$. Dann existiert ein $C'' \in \text{Ch}(\varphi(X))$ mit $C'' \notin \mathcal{P}$.

Wir zeigen, daß $\mathcal{R}' := (\mathcal{R} \setminus C'') \setminus \{\emptyset\}$ ein Reconciled Tree in $\text{RC}(\mathcal{G}, \mathcal{S})$ ist. $T_{\mathcal{R}}(C'')$ ist nicht im Wertebereich von φ enthalten, womit $\varphi' : \mathcal{G} \rightarrow \mathcal{R}'$ mit $\varphi'(Y) \mapsto \varphi(Y) \setminus C''$ ein minimaler spezieskonformer Einbettungsmonomorphismus ist. Es gilt $B(\varphi(X)) = B(C'')$, da $\varphi(X)$ ein Duplikationselement in einem m-Clusterelement ist. Also gilt $B(\mathcal{R}') = B(\mathcal{R})$, womit \mathcal{R}' von \mathcal{S} ableitbar ist. $(\mathcal{R} \setminus C'') \setminus \{\emptyset\}$ ist ein m-Clusterelement, bis auf die Möglichkeit, daß $\varphi'(X)$ nicht mehr mindestens zwei Kinder besitzen könnte. Da aber φ nicht nach $T_{\mathcal{R}}(C'')$ abbildet, besitzt $\varphi'(X)$ mindestens zwei Kinder, womit \mathcal{R}' ein m-Clusterelement ist. Somit ist $\mathcal{R}' \in \text{RC}(\mathcal{G}, \mathcal{S})$ und $|\mathcal{R}'| < |\mathcal{R}|$, was ein Widerspruch zur Minimalität von $|\mathcal{R}|$ ist. \square

Verw. von: Lem. 4.11, 4.6; Verw. in: Theo. 4.8, Lem. 4.17, 5.1.

Wir werden nun für ein festes aber beliebiges $Z \in \text{Ch}(X)$ die Topologie von $e_{\mathcal{R}}(Z)$ beschreiben, indem wir $B_{e_{\mathcal{R}}(Z)} \in \text{Oiso}(e_{\mathcal{R}}(Z), T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \setminus T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z)))$ zeigen. Dazu skizzieren wir zunächst die Beweisidee und zeigen dann den formalen Beweis hierzu. Die hier gezeigte Beweisidee ist analog zu der Beweisidee von Theorem 4.5.

Grundlage des Beweises ist die Konstruktion eines Reconciled Tree $\mathcal{R}'' \in \text{RC}(\mathcal{G}, \mathcal{S})$, in welchen \mathcal{G} mit dem minimalen Einbettungsmonomorphismus φ'' spezieskonform einbettbar ist.⁴ \mathcal{R}'' besitzt die beiden folgenden Eigenschaften: $|e_{\mathcal{R}''}(Z)| = |T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \setminus T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z))|$ und $|e_{\mathcal{R}}(Z)| = |e_{\mathcal{R}''}(Z)|$. Hieraus folgt die wesentliche Aussage

$$|e_{\mathcal{R}}(Z)| = |T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \setminus T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z))|.$$

Wir wollen $B_{e_{\mathcal{R}}(Z)} \in \text{Oiso}(e_{\mathcal{R}}(Z), T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \setminus T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z)))$ zeigen. Lemma 4.8 besagt, daß $B_{e_{\mathcal{R}}(Z)} \in \text{Ehom}(e_{\mathcal{R}}(Z), \mathcal{S})$ und damit $B_{e_{\mathcal{R}}(Z)} \in \text{Ohom}(e_{\mathcal{R}}(Z), \mathcal{S})$

⁴ $e_{\mathcal{R}''}$ ist damit die vereinfachte Form von $e_{\mathcal{R}'', \varphi''}$.

gilt. Also gilt auch $B_{e_{\mathcal{R}}(Z)} \in \text{Oepi}(e_{\mathcal{R}}(Z), B(e_{\mathcal{R}}(Z)))$. Die Aussage $|e_{\mathcal{R}}(Z)| = |\mathcal{T}_S(\bigsqcup_S(X)) \setminus \mathcal{T}_S(\bigsqcup_S(Z))|$ verwenden wir nun einerseits, um $B(e_{\mathcal{R}}(X)) = \mathcal{T}_S(\bigsqcup_S(X)) \setminus \mathcal{T}_S(\bigsqcup_S(Z))$ zu zeigen, womit $B_{e_{\mathcal{R}}(Z)} \in \text{Oepi}(e_{\mathcal{R}}(Z), \mathcal{T}_S(\bigsqcup_S(X)) \setminus \mathcal{T}_S(\bigsqcup_S(Z)))$ folgt. Andererseits folgt mit $|e_{\mathcal{R}}(Z)| = |\mathcal{T}_S(\bigsqcup_S(X)) \setminus \mathcal{T}_S(\bigsqcup_S(Z))|$ nun die zu zeigende Aussage $B_{e_{\mathcal{R}}(Z)} \in \text{Oiso}(e_{\mathcal{R}}(Z), \mathcal{T}_S(\bigsqcup_S(X)) \setminus \mathcal{T}_S(\bigsqcup_S(Z)))$.

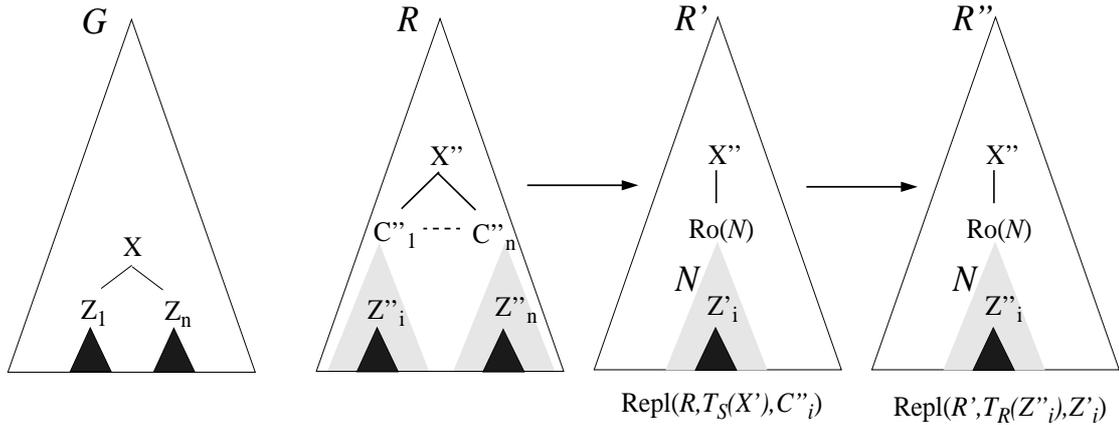


Abbildung 4.11: Die Konstruktion von \mathcal{R}''

Konstruieren wir nun \mathcal{R}' mit Hilfe der Operationen Repl und \times . Abbildung 4.11 veranschaulicht die Konstruktion von \mathcal{R}'' . Das Element $X \in \mathcal{G}$ wird durch φ auf $X'' := \varphi(X)$ und dessen Kinder $Z_1, \dots, Z_n \in \text{Ch}(X)$ auf die Elemente $Z'_1 := \varphi(Z_1), \dots, Z'_n := \varphi(Z_n)$ in \mathcal{R} abgebildet. Die Kinder von X'' werden durch C''_1, \dots, C''_n bezeichnet. Nach Theorem 4.7 ist $E_{\mathcal{R}}(X)$ eine Partition aus $\varphi(X)$ und allen lokalen Entwicklungsumgebungen von $E_{\mathcal{R}}(X)$. In Abbildung 4.11 ist C''_i für ein $i \in \{1, \dots, n\}$ das Maximum der lokalen Entwicklungsumgebung $e_{\mathcal{R}}(Z_i)$.

Da $Z \in \text{Ch}(X)$ ist, existiert ein $i \in \{1, \dots, n\}$ mit $Z_i = Z$. Um \mathcal{R}'' zu konstruieren, ersetzen wir zuerst $\mathcal{T}_{\mathcal{R}}(C''_i)$ durch $\mathcal{N} := \mathcal{T}_S(B(C''_i)) \times q$, für ein zu $\text{Ro}(\mathcal{R})$ tdisjunktes $p \in \mathbb{N}$. Wir erhalten damit das Mengensystem \mathcal{R}' , welches aus \mathcal{S} ableitbar ist.

Wir zeigen $\mathcal{N} = \mathcal{T}_S(\bigsqcup_S(X)) \times q$. Nach dem Duplikations-Theorem 4.6 ist X'' ein Duplikationselement, womit $B(X'') = B(C''_i)$ folgt. Aus dem LCA-Theorem 4.4 folgt $B(X'') = \bigsqcup_S(X)$ und damit gilt $\bigsqcup_S(X) = B(C''_i)$. Also gilt $\mathcal{N} = \mathcal{T}_S(B(C''_i)) \times q = \mathcal{T}_S(\bigsqcup_S(X)) \times q$.

Somit haben wir $\mathcal{T}_{\mathcal{R}}(C''_i)$ durch die duplikationsfreie Ableitung $\mathcal{T}_S(\bigsqcup_S(X)) \times q$ von $\mathcal{T}_S(\bigsqcup_S(X))$ ersetzt, wobei \mathcal{R}' von \mathcal{S} ableitbar ist. Um die spezieskonforme Einbettung von \mathcal{G} sicherzustellen, ersetzen wir in \mathcal{R}' den Teilbaum gewurzelt an $Z'_i := \bigsqcup_S(Z_i) \times \{q\}$ durch $\mathcal{T}_{\mathcal{R}}(Z'_i)$ und erhalten den Reconciled Tree \mathcal{R}'' .

Für \mathcal{R}'' gilt: Die Topologien von $\mathcal{R}'' \setminus e_{\mathcal{R}''}(Z)$ und $\mathcal{R} \setminus e_{\mathcal{R}}(Z)$ sind gleich und $|e_{\mathcal{R}''}(Z)| = |\mathsf{T}_s(\bigsqcup_s(X)) \setminus \mathsf{T}_s(\bigsqcup_s(Z))|$. Aus der Gleichheit der Topologien von $\mathcal{R}'' \setminus e_{\mathcal{R}''}(Z)$ und $\mathcal{R} \setminus e_{\mathcal{R}}(Z)$ folgt $|\mathcal{R}'' \setminus e_{\mathcal{R}''}(Z)| = |\mathcal{R} \setminus e_{\mathcal{R}}(X)|$. Da $\mathcal{R}, \mathcal{R}'' \in \mathsf{RC}(\mathcal{G}, \mathcal{S})$ gilt, folgt $|\mathcal{R}| = |\mathcal{R}''|$. Aus $|\mathcal{R}'' \setminus e_{\mathcal{R}''}(Z)| = |\mathcal{R} \setminus e_{\mathcal{R}}(X)|$ und $|\mathcal{R}| = |\mathcal{R}'|$ folgt dann $|e_{\mathcal{R}}(Z)| = |e_{\mathcal{R}''}(Z)|$. Somit erfüllt \mathcal{R}'' die geforderten Eigenschaften $|e_{\mathcal{R}''}(Z)| = |\mathsf{T}_s(\bigsqcup_s(X)) \setminus \mathsf{T}_s(\bigsqcup_s(Z))|$ und $|e_{\mathcal{R}}(Z)| = |e_{\mathcal{R}''}(Z)|$.

Theorem 4.8

Seien $X \in \mathcal{G}$, $Z \in \mathsf{Ch}(X)$ und $\varphi(X)$ ein Duplikationselement. Dann gilt

$$\mathsf{B}_{e_{\mathcal{R}}(Z)} \in (e_{\mathcal{R}}(Z), \mathsf{T}_s(\bigsqcup_s(X)) \setminus \mathsf{T}_s(\bigsqcup_s(Z))).$$

Beweis. Der Beweis ist analog zu dem Beweis von Theorem 4.5. Zur Vereinfachung definieren wir $\mathcal{P} := \mathsf{T}_s(\bigsqcup_s(C'')) \setminus \mathsf{T}_s(\bigsqcup_s(Z))$. Nach Theorem 4.7 gilt $\mathsf{E}_{\mathcal{R}}(X) = \{\varphi(X)\} \dot{\cup} \bigcup_{K \in \mathsf{Ch}(X)} e_{\mathcal{R}}(K)$, womit genau ein Kind $C'' \in \mathsf{Ch}(X)$ mit $e_{\mathcal{R}}(Z) = \mathsf{T}_{\mathcal{R}}(C'') \setminus \mathsf{T}_{\mathcal{R}}(\varphi(Z))$ existiert. Wir zeigen zuerst, daß $|e_{\mathcal{R}}(Z)| = |\mathcal{P}|$ gilt. Dazu definieren wir, basierend auf \mathcal{R} , das Multimengensystem \mathcal{R}'' für ein zu $\mathsf{Ro}(\mathcal{R})$ tdisjunktes $p \in \mathbb{N}$.

$$\mathcal{R}' := \mathsf{Repl}(\mathcal{R}, \mathsf{T}_s(\mathsf{B}(C'')) \times p, C'')$$

$$\mathcal{R}'' := \mathsf{Repl}(\mathcal{R}', \mathsf{T}_{\mathcal{R}}(\varphi(Z)), \mathsf{B}(\varphi(Z)) \times \{p\})$$

C'' ist das Maximum von $e_{\mathcal{R}}(Z)$, womit Z das einzige Kind aus $\mathsf{Ch}(X)$ mit $\varphi(Z) \subseteq C''$ ist. Damit ist Lemma 9.3 anwendbar, woraus $\mathcal{R}'' \in \mathsf{RC}(\mathcal{G}, \mathcal{S})$ folgt. Die Kardinalität von \mathcal{R}'' berechnet sich anhand der Kardinalität von \mathcal{R} wie folgt:

$$\begin{aligned} |\mathcal{R}''| &= |\mathcal{R}| - |\mathsf{T}_{\mathcal{R}}(C'')| + |\mathsf{T}_s(\mathsf{B}(C'')) \times p| \\ &\quad - |\mathsf{T}_s(\mathsf{B}(\varphi(Z))) \times \{p\}| + |\mathsf{T}_{\mathcal{R}}(\varphi(Z))| \\ &= |\mathcal{R}| - (|\mathsf{T}_{\mathcal{R}}(C'')| - |\mathsf{T}_{\mathcal{R}}(\varphi(Z))|) \\ &\quad + |\mathsf{T}_s(\mathsf{B}(C''))| - |\mathsf{T}_s(\mathsf{B}(\varphi(Z)))| \\ &= |\mathcal{R}| - (|\mathsf{T}_{\mathcal{R}}(C'')| - |\mathsf{T}_{\mathcal{R}}(\varphi(Z))|) \\ &\quad + |\mathsf{T}_s(\mathsf{B}(\varphi(X)))| - |\mathsf{T}_s(\mathsf{B}(\varphi(Z)))| \end{aligned} \tag{4.17}$$

$$\begin{aligned} &= |\mathcal{R}| - (|\mathsf{T}_{\mathcal{R}}(C'')| - |\mathsf{T}_{\mathcal{R}}(\varphi(Z))|) \\ &\quad + |\mathsf{T}_s(\bigsqcup_s(X))| - |\mathsf{T}_s(\bigsqcup_s(Z))| \end{aligned} \tag{4.18}$$

$$\begin{aligned} &= |\mathcal{R}| - (|\mathsf{T}_{\mathcal{R}}(C'') \setminus \mathsf{T}_{\mathcal{R}}(\varphi(Z))|) \\ &\quad + |\mathsf{T}_s(\bigsqcup_s(X)) \setminus \mathsf{T}_s(\bigsqcup_s(Z))| \end{aligned} \tag{4.19}$$

$$= |\mathcal{R}| - |e_{\mathcal{R}}(Z)| + |\mathcal{P}|.$$

Zu (4.17): Da $C'' \in \text{Ch}(X)$ gilt und $\varphi(X)$ ein Duplikationselement ist, folgt in einem m -Clustersystem $B(C'') = B(\varphi(X))$.

Zu (4.18): Folgt aus dem LCA-Theorem 4.4.

Zu (4.19): Die Umrechnung in $T_{\mathcal{R}}(C'')$ gilt aufgrund der Einbettungsmonomorphie von φ , die Umrechnung in $T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X))$ gilt, da $\bigsqcup_{\mathcal{S}}$ ordnungshomomorph ist.

Da $\mathcal{R}, \mathcal{R}^n \in \text{RC}(\mathcal{G}, \mathcal{S})$ gilt, folgt $|\mathcal{R}| = |\mathcal{R}^n|$ (aufgrund der Minimalität beider m -Clustersysteme). Somit folgt $|e_{\mathcal{R}}(Z)| = |\mathcal{P}|$.

Da $e_{\mathcal{R}}(X)$ gegen \sqcup abgeschlossen ist, folgt $B_{e_{\mathcal{R}}(Z)} \in \text{Ehom}(e_{\mathcal{R}}(X), \mathcal{S})$ aus Lemma 4.8. Daraus folgt $B_{e_{\mathcal{R}}(Z)} \in \text{Ohom}(e_{\mathcal{R}}(Z), \mathcal{S})$, und es gilt

$$B_{e_{\mathcal{R}}(Z)} \in \text{Oepi}(e_{\mathcal{R}}(Z), B_{e_{\mathcal{R}}(Z)}(e_{\mathcal{R}}(Z))).$$

Wir zeigen, daß $B(e_{\mathcal{R}}(Z)) = \mathcal{P}$ gilt. $B(e_{\mathcal{R}}(Z)) \supseteq \mathcal{P}$ gilt nach der folgenden Aussage:

$$\begin{aligned} B(e_{\mathcal{R}}(Z)) &= B(T_{\mathcal{R}}(C'') \setminus T_{\mathcal{R}}(\varphi(Z))) \\ &\supseteq B(T_{\mathcal{R}}(C'')) \setminus B(T_{\mathcal{R}}(\varphi(Z))) && \text{Lemma 3.3} \\ &= T_{\mathcal{S}}(B(C'')) \setminus T_{\mathcal{S}}(B(\varphi(Z))) && \mathcal{R} \text{ ist aus } \mathcal{S} \text{ ableitbar.} \end{aligned}$$

Da $\varphi(X) \in \text{Dup}_{\mathcal{R}}$ und $C'' \in \text{Ch}(\varphi(X))$ gilt, folgt $B(C'') = B(\varphi(X))$, und es gilt weiter:

$$\begin{aligned} &= T_{\mathcal{S}}(B(\varphi(X))) \setminus T_{\mathcal{S}}(B(\varphi(Z))) \\ &= T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \setminus T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z)) && \text{LCA-Theorem 4.4} \\ &= \mathcal{P}. \end{aligned}$$

Daraus folgt $|B(e_{\mathcal{R}}(Z))| \geq |\mathcal{P}|$. Es gilt $|e_{\mathcal{R}}(Z)| \geq |B(e_{\mathcal{R}}(Z))|$, und wir haben gerade $|e_{\mathcal{R}}(Z)| = |\mathcal{P}|$ gezeigt. Also gilt

$$|e_{\mathcal{R}}(Z)| \geq |B(e_{\mathcal{R}}(Z))| \geq |\mathcal{P}| = |e_{\mathcal{R}}(Z)|,$$

woraus $|B(e_{\mathcal{R}}(Z))| = |\mathcal{P}|$ folgt. Mit $B(e_{\mathcal{R}}(Z)) \supseteq \mathcal{P}$ folgt hieraus $B(e_{\mathcal{R}}(Z)) = \mathcal{P}$.

Somit ist $B_{e_{\mathcal{R}}(Z)} \in \text{Oepi}(e_{\mathcal{R}}(Z), \mathcal{P})$, und da $|e_{\mathcal{R}}(Z)| = |\mathcal{P}|$ gilt, folgt $B_{e_{\mathcal{R}}(Z)} \in \text{Oiso}(e_{\mathcal{R}}(Z), \mathcal{P})$. \square

Verw. von: Lem. 4.8, 3.3, Theo. 4.4, 4.7; Verw. in: .

Wir werden nun mit Kenntnis der Topologie von $E_{\mathcal{R}}(X)$ das Duplikationstheorem 4.6 durch Duplikations-Korollar 4.5 verschärfen. Dazu zeigen wir zunächst das folgende Lemma und dann den Duplikations-Korollar.

Lemma 4.17

Sei $X \in \mathcal{G}$, und $E_{\mathcal{R}}(X)$ besitzt ein Duplikationselement. Dann ist $\varphi(X)$ das einzige Duplikationselement in $E_{\mathcal{R}}(X)$.

Beweis. Aus dem Duplikations-Theorem 4.6 folgt, daß $\varphi(X)$ ein Duplikationselement ist, da $E_{\mathcal{R}}(X)$ ein Duplikationselement besitzt. Nach Theorem 4.7 ist $E_{\mathcal{R}}(X)$ eine Partition aus $\varphi(X)$ und allen seinen lokalen Entwicklungsumgebungen. Wir zeigen, daß eine feste aber beliebige lokale Entwicklungsumgebung $e_{\mathcal{R}}(Z)$ mit $Z \in \text{Ch}(X)$ keine Duplikationselemente besitzt, womit $\varphi(X)$ dann das einzige Duplikationselement in $E_{\mathcal{R}}(X)$ ist.

Nach Theorem 4.8 gilt, $B_{e_{\mathcal{R}}(Z)} \in \text{Oiso}(e_{\mathcal{R}}(Z), T_s(\bigsqcup_S(Z) \setminus T_s(\bigsqcup_S(Z))))$, woraus $|B(e_{\mathcal{R}}(Z))| = |e_{\mathcal{R}}(Z)|$ folgt. Aus Lemma 4.12 folgt hieraus, daß $e_{\mathcal{R}}(Z)$ kein Duplikationselement beinhaltet. \square

Verw. von: Theo. 4.6, 4.7, 4.8, Lem. 4.12; Verw. in: Kor. 4.5.

Korollar 4.5 (Duplikations-Korollar)

Sei $X \in \mathcal{R}$.

$E_{\mathcal{R}}(X)$ besitzt kein Duplikationselement, wenn $P_{\text{NoDup}}(X)$ gilt.

$E_{\mathcal{R}}(X)$ besitzt nur das Duplikationselement $\varphi(X)$, wenn $P_{\text{NoDup}}(X)$ nicht gilt.

Beweis. Folgt aus dem Duplikations-Theorem 4.6 und Lemma 4.17. \square

Verw. von: Theo. 4.6, Lem. 4.17; Verw. in: Kor. 5.1.

Kapitel 5

Konstruktive Definitionen zum Reconciled Tree

Die Definition des Reconciled Tree ist direkt aus dem biologischen Sachverhalt heraus formuliert und dadurch gut wiedergegeben worden, Dazu hat Page einen Reconciled Tree als ein an Kardinalität minimales Element einer Menge von m -Clustersystemen mit gewissen Eigenschaften beschrieben. Eine konkrete Vorstellung der Topologie des Reconciled Tree fehlt somit. Diese nicht konstruktive Definition erschwert die biologische Interpretation eines Reconciled Tree, wie auch dessen Konstruktion und den formalen Umgang mit diesem. Entsprechend ist auch das Duplikationselement zwar biologisch einsichtig in die Definition übertragen, jedoch nicht konstruktiv definiert worden.

In dem Abschnitt 4.4 haben wir Eigenschaften des Reconciled Tree und des Duplikationselementes gezeigt, mit denen wir diese konstruktiv definieren können. In Abschnitt 5.1 werden wir zunächst das Duplikationselement und dann in Abschnitt 5.2 den Reconciled Tree konstruktiv definieren. Dazu betrachten wir einen Gen- und Speziesbaum $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$ über einer Menge M und einen Reconciled Tree $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$, in welchen \mathcal{G} durch einen minimalen Einbettungsmonomorphismus φ spezieskonform einbettbar ist.

5.1 Duplikationselemente in \mathcal{G}

Ein Duplikationselement ist bisher anhand eines Reconciled Tree und damit nicht konstruktiv definiert worden. In diesem Abschnitt werden wir das Duplikationselement konstruktiv, also ohne die direkte Verwendung eines Reconciled Tree, aus $\text{RC}(\mathcal{G}, \mathcal{S})$ definieren.

Nach dem Duplikations-Korollar 4.5 ist nur ein Element $\varphi(X)$ ein Duplikationselement in \mathcal{R} , für welches das Prädikat $P_{\text{NoDup}}(X)$ nicht gilt. Die Elemente $\varphi(X)$ und X bezeichnen, jeweils in \mathcal{R} und \mathcal{G} , dasselbe Gen. Somit können wir ein Duplikations-Gen, welches durch ein Duplikationselement in \mathcal{R} bezeichnet ist, auch durch ein *Duplikationselement in \mathcal{G}* bezeichnen.

Definition 5.1 (Duplikationselement im Genbaum)

Sei M eine Menge, $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$ ein Gen- und Speziesbaum. Dann bezeichnen wir mit

$$\text{DupK}_{\mathcal{G}, \mathcal{S}} := \{X \in \mathcal{G} \mid \neg P_{\text{NoDup}}(X)\}$$

die Menge aller Duplikationselemente in \mathcal{G} . Wir bezeichnen mit $\overline{\text{DupK}}_{\mathcal{G}, \mathcal{S}} := \mathcal{G} \setminus \text{DupK}_{\mathcal{G}, \mathcal{S}}$ die Elemente in \mathcal{G} , welche nicht Duplikationselemente sind.

Sind \mathcal{G}, \mathcal{S} offensichtlich, so verwenden wir die Bezeichnungen DupK oder $\overline{\text{DupK}}$.

Korollar 5.1

Sei M eine Menge, $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$, $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$ und φ ein minimaler Einbettungsmonomorphismus von \mathcal{G} in \mathcal{R} , dann gilt

$$\varphi(\text{DupK}_{\mathcal{G}, \mathcal{S}}) = \text{Dup}_{\mathcal{R}}.$$

Beweis. Folgt direkt aus dem Duplikations-Korollar 4.5. □

Verw. von: Kor. 4.5; *Verw. in:* .

Mit dieser konstruktiven Definition des Duplikationselementes werden wir in dem folgenden Abschnitt den Reconciled Tree konstruktiv definieren.

5.2 Der Reconciled Tree, konstruktiv definiert

In Abschnitt 4.4 haben wir die Topologie eines Reconciled Tree beschrieben. Alle Reconciled Trees in $\text{RC}(\mathcal{G}, \mathcal{S})$ besitzen eine eindeutige Topologie, welche sich aus Subtopologien von \mathcal{G} und \mathcal{S} kombiniert. Diese Kombination ist wesentlich für die biologische Auswertung eines Reconciled Tree. Die Bezeichnungen der Elemente eines Reconciled Tree geben diese Kombination jedoch nur indirekt, durch das Multimengenkonzept, wieder. Wir werden daher den Reconciled Tree konstruktiv definieren, indem wir den Elementen von \mathcal{R} Bezeichnungen zuordnen, welche direkt auf die Subtopologien von \mathcal{G} und \mathcal{S} verweisen und dadurch den Reconciled Tree konkret beschreiben.

Ein Beispiel des konstruktiv definierten Reconciled Tree $RK_{\mathcal{G},\mathcal{S}}$ von \mathcal{G} über \mathcal{S} zeigt Abbildung 5.1. Die Mengen a, b, c und A, B, C bestehen dabei aus den Blättern der durch sie definierten Teilbäume, wie z.B. $B = \{2, 3, 4, 5\}$. Ein Dreitupel $(G_{\text{from}}, G_{\text{to}}, S_{\text{host}})$ von $RK_{\mathcal{G},\mathcal{S}}$ beschreibt dabei ein Gen, welches während der Entwicklung des Gens $G_{\text{from}} \in \mathcal{G}$ in das Gen $G_{\text{to}} \in \mathcal{G}$ in der Host-Spezies $S_{\text{host}} \in \mathcal{S}$ entstanden ist. Betrachten wir das Dreitupel $(b, 2, B)$. Das Gen b ist in der Host-Spezies A , das Gen 2 in der Host-Spezies B beinhaltet. Damit hat sich das Gen b in der Host-Spezies A über das Gen $(b, 2, B)$ in der Host-Spezies B zum Gen 2 in der Host-Spezies B entwickelt.

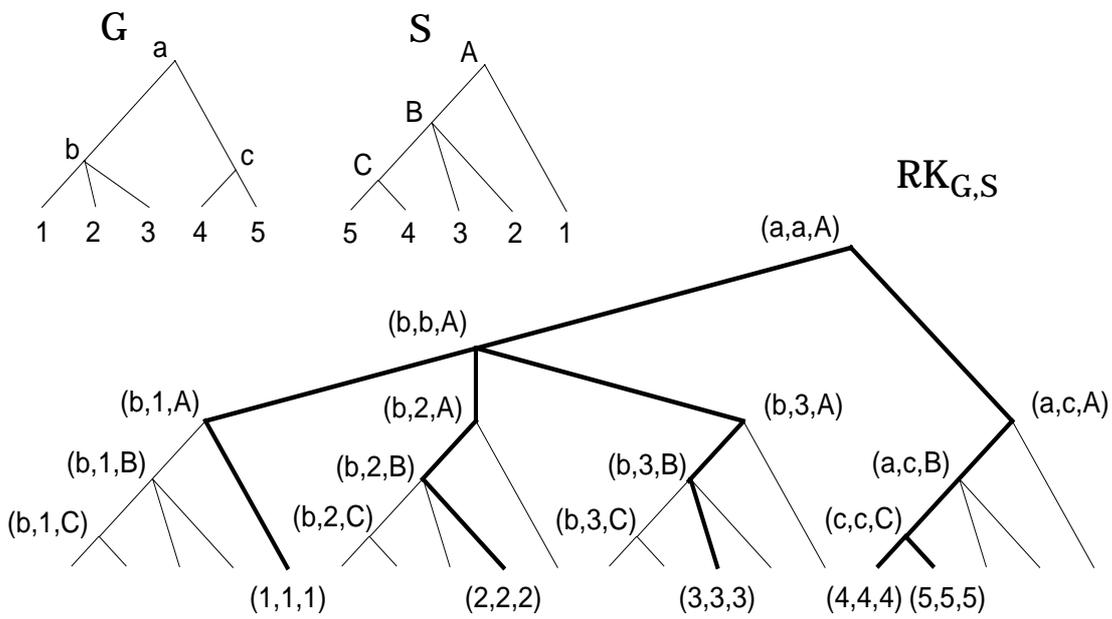


Abbildung 5.1: Der konstruktive Reconciled Tree $RK_{\mathcal{G},\mathcal{S}}$

Diese konstruktive Definition des Reconciled Tree werden wir in Abschnitt 5.2.1 aus dem Reconciled Tree von R. Page herleiten. In Abschnitt 5.2.2 formulieren wir dann formal den konstruktiven Reconciled Tree und zeigen, daß dieser dieselbe Topologie wie ein nicht konstruktiv definierter Reconciled Tree besitzt.

5.2.1 Der konstruktive Reconciled Tree: Die Idee

In diesem Abschnitt werden wir allen Elementen in \mathcal{R} biologisch interpretierbare Dreitupel-Bezeichnungen zuordnen, welche sich auch formal einfach handhaben lassen. Ein Element $Y'' \in \mathcal{R}$ bezeichnen wir dabei durch das Dreitupel (X, Z, S) mit $X \in \mathcal{G}$, $Z \in \text{Ch}(X)$ und $S \in \mathcal{S}$. X und Z geben dabei die *Ordnungs-Position* von Y'' in \mathcal{R} bezüglich der Topologie von \mathcal{G} an. S gibt die *Ordnungs-Position*

von Y'' bezüglich der Topologie von \mathcal{S} an. Die genaue Bedeutung der Parameter einer Dreitupel-Bezeichnung beschreiben wir in den folgenden drei Abschnitten.

5.2.1.1 Zuordnung des ersten Dreitupelparameters

Betrachten wir zunächst den Bezug eines Elementes in \mathcal{R} zu der Topologie von \mathcal{G} . Durch den Einbettungsmonomorphismus φ überträgt sich die Topologie von \mathcal{G} auf \mathcal{R} . Diese Topologie partitioniert \mathcal{R} in Regionen, die eine von \mathcal{G} unabhängige Topologie besitzen. Eine solche Region wird durch die Elemente in \mathcal{R} , welche "zwischen" den Bildern eines Elementes $V \in \mathcal{G}$ und den Bildern seiner Kindern $\text{Ch}(V)$ liegen, gebildet. In Definition 4.9 haben wir diese Regionen als Entwicklungsumgebung von V mit $E_{\mathcal{R}}(V) = T_{\mathcal{R}}(\varphi(V)) \setminus \bigcup_{Z \in \text{Ch}(V)} T_{\mathcal{R}}(\varphi(Z))$ definiert, wobei $\mathcal{R} = \bigcup_{V \in \mathcal{G}} E_{\mathcal{R}}(V)$ nach Lemma 4.7 gilt. Jedem Element in \mathcal{R} kann damit eindeutig eine Entwicklungsumgebung in \mathcal{R} zugewiesen werden. Das Element Y'' liegt also in einer Entwicklungsumgebung $E_{\mathcal{R}}(X)$, welches durch die erste Position des Dreitupels (X, Z, S) gekennzeichnet wird. Abbildung 5.2 zeigt an dem Reconciled Tree aus Abbildung 5.1 die Zuordnung des jeweils ersten Parameters jedes Dreitupels.

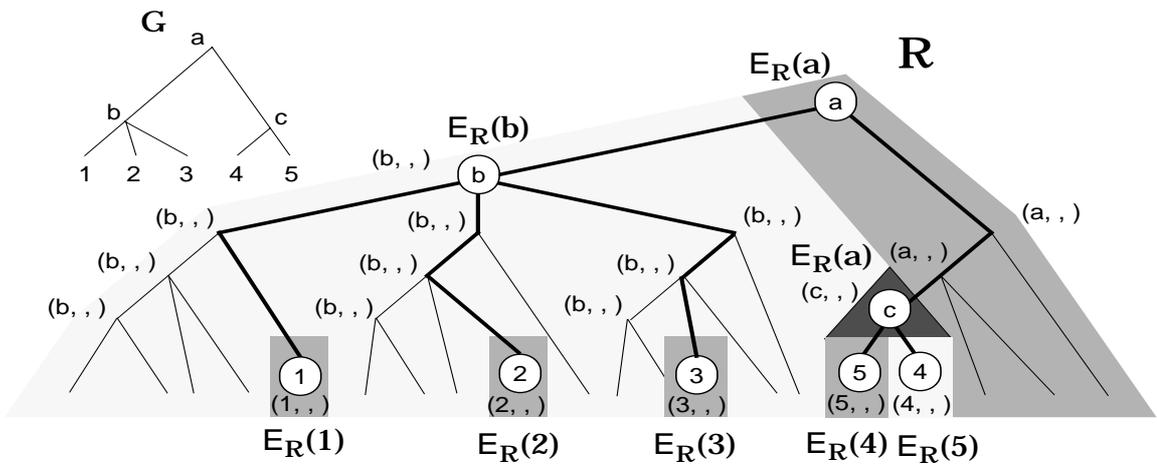


Abbildung 5.2: Der konstruktive Reconciled Tree $\text{RK}(\mathcal{G}, \mathcal{S})$

Betrachten wir die biologische Bedeutung der Entwicklungsumgebung $E_{\mathcal{R}}(X)$. In \mathcal{G} hat sich das Gen X zu allen seinen direkten Nachfolge-Genen in $\text{Ch}(X)$ entwickelt. Diese Gene tragen in \mathcal{R} die Bezeichnungen des Elementes $\varphi(X)$ und der Elemente aus der Menge $\varphi(\text{Ch}(X))$. Alle Gene, die in \mathcal{R} durch die Entwicklung von $\varphi(X)$ zu seinen Kindern $\varphi(\text{Ch}(X))$ entstanden, sind in \mathcal{G} nicht bekannt.¹

¹Diese in \mathcal{R} unbekanntenen Gene werden in \mathcal{G} durch die Kanten dargestellt.

Wir bezeichnen allgemein Gene in \mathcal{R} , welche in \mathcal{G} nicht bekannt sind, als *unbekannte Gene*. Damit ist $\mathcal{R} \setminus \varphi(\mathcal{G})$ die Menge aller unbekannter Gene in \mathcal{R} . Entsprechend bezeichnen wir die Gene in $\varphi(\mathcal{G})$ als *bekannte Gene*.

Die Entwicklungsumgebung $E_{\mathcal{R}}(X)$ ohne $\varphi(X)$ beschreibt genau die unbekannt Genen in \mathcal{R} , welche durch die Entwicklung von X in \mathcal{R} ($\varphi(X)$) zu seinen Kindern $\text{Ch}(X)$ in \mathcal{R} ($\varphi(\text{Ch}(X))$) entstanden sind. Ein Gen, welches in \mathcal{R} mit (X, \cdot) gekennzeichnet ist, beschreibt somit entweder das Gen $\varphi(X)$ oder ein unbekanntes Gen, das durch die Entwicklung von X zu einem Gen $Z \in \text{Ch}(X)$ in \mathcal{R} entstanden ist.

5.2.1.2 Zuordnung des zweiten Dreitupelparameters

Wir haben jedem Element aus \mathcal{R} mit dem ersten Parameter seiner Dreitupelbezeichnung eindeutig die Entwicklungsumgebung zugeordnet, welche dieses Element beinhaltet. Mit dem zweiten Parameter werden wir ein Element eindeutig dessen lokaler Entwicklungsumgebung, dem Element $\varphi(\mathcal{G})$ oder dessen unbestimmter Entwicklungsumgebung zuordnen.

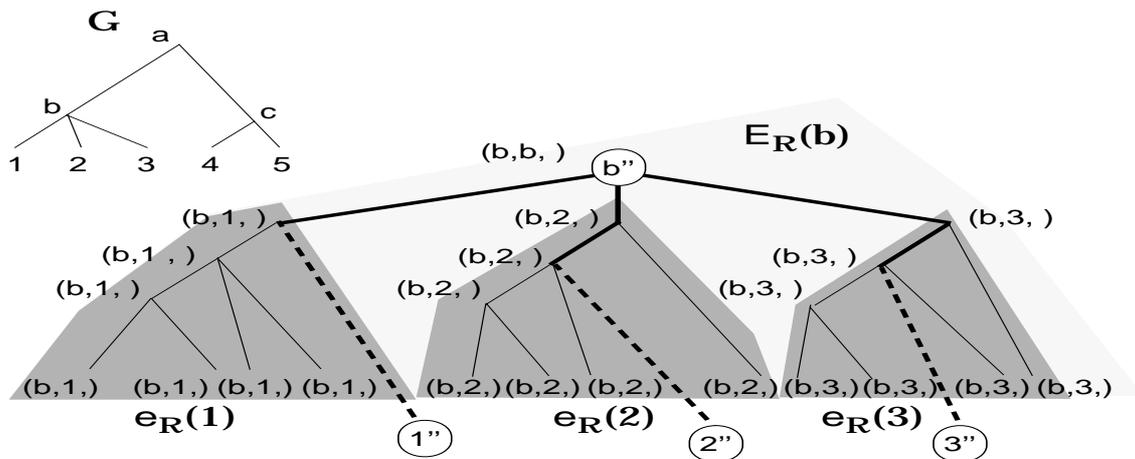


Abbildung 5.3: Die Dreitupelbezeichnungen von $E_{\mathcal{R}}(b)$

Betrachten wir die Entwicklungsumgebung $E_{\mathcal{R}}(X)$ in \mathcal{R} . Nach Lemma 4.11 lässt sich $E_{\mathcal{R}}(X)$ in alle seine lokalen Entwicklungsumgebungen $e_{\mathcal{R}}(Z)$ mit $Z \in \text{Ch}(X)$, die Menge $\{\varphi(X)\}$ und dessen unbestimmte Entwicklungsumgebung $\tilde{E}_{\mathcal{R}}(X)$ partitionieren. Jedem Element in $E_{\mathcal{R}}(X)$ ist damit eindeutig eine dieser Mengen zuzuordnen. Einem Element der lokalen Entwicklungsumgebung $e_{\mathcal{R}}(Z)$ mit $Z \in \text{Ch}(X)$ ordnen wir das Dreitupel (X, Z, \cdot) zu, dem Element $\varphi(X)$ ordnen wir

das Dreitupel $(X, X,)$ zu und einem Element der unbekanntem Entwicklungsumgebung $\tilde{E}_{\mathcal{R}}(X)$ ordnen wir das Dreitupel $(X, ?,)$ zu.

Die Abbildung 5.3 zeigt die Zuordnung des zweiten Parameters der Dreitupelbezeichnung in $E_{\mathcal{R}}(b)$. Das Theorem 4.7 besagt, daß, falls X ein Duplikationselement ist, die Entwicklungsumgebung $\varphi(X)$ keine unbestimmte Entwicklungsumgebung $\tilde{E}_{\mathcal{R}}(X)$ besitzt. Da b ein Duplikationselement ist, beinhaltet $E_{\mathcal{R}}(b)$ somit keine Elemente mit der Dreitupelbezeichnung $(X, ?,)$.

Biologisch interpretiert, beschreibt eine lokale Entwicklungsumgebung $e_{\mathcal{R}}(Z)$ mit $Z \in \text{Ch}(X)$ alle unbekanntem Gene in \mathcal{R} , welche durch die Entwicklung von X zu Z in \mathcal{R} entstanden sind. Ein Gen in \mathcal{R} mit der Dreitupelbezeichnung $(X, Z,)$ beschreibt somit ein unbekanntem Gen, welches durch die Entwicklung von X zu seinem Kind Z in \mathcal{R} entstanden ist. Ist ein Gen in \mathcal{R} mit $(X, X,)$ bezeichnet, so hat sich dieses Gen bei der Entwicklung von X nach X in \mathcal{R} entwickelt und ist somit das bekannte Gen $\varphi(X)$. Damit ist ein durch (b_1, b_2, b_3) bezeichnetes Gen $Y'' \in \mathcal{R}$ genau dann bekannt, wenn $b_1 = b_2$ gilt. Ist Y'' bekannt, so gilt $Y'' = \varphi(b_1)$.

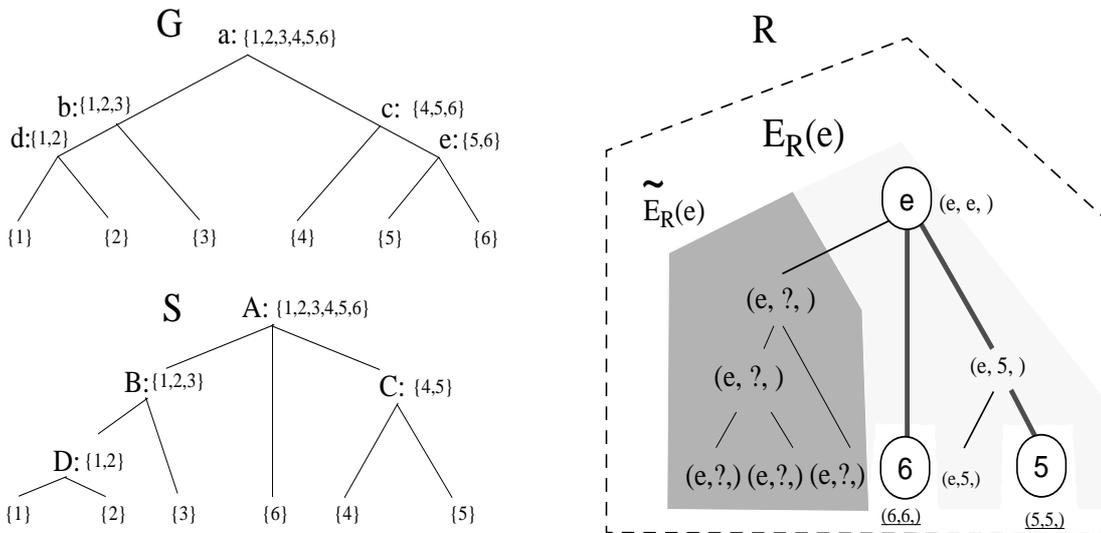


Abbildung 5.4: Die Dreitupelbezeichnungen von $\tilde{E}_{\mathcal{R}}(e)$

Betrachten wir die unbestimmte Entwicklungsumgebung $\tilde{E}_{\mathcal{R}}(e)$ in Abbildung 5.4. Die Gene dieser Umgebung sind entweder durch die Entwicklung des Gens e zu dem Gen 5 oder dem Gen 6 entstanden. Die Multifurkation läßt also eine genauere biologische Interpretation der Entwicklung der Gene aus der unbekanntem Entwicklungsumgebung nicht zu. Dies drücken wir durch das Symbol “?” aus.

5.2.1.3 Zuordnung des dritten Dreitupelparameters

Mit den ersten beiden Parametern der Dreitupel-Bezeichnung eines Gens $Y'' \in \mathcal{R}(X)$ haben wir dessen Bezug zu dem Genbaum \mathcal{G} beschrieben. Dadurch haben wir Y'' eindeutig einer lokalen Entwicklungsumgebung $e_{\mathcal{R}}(Z)$ für ein $Z \in \text{Ch}(X)$, der unbestimmten Entwicklungsumgebung $\tilde{E}_{\mathcal{R}}(X)$ oder dem bekannten Gen $\varphi(X)$ zugeordnet. In jeder dieser Umgebungen beschreiben wir nun das Element Y'' eindeutig durch den dritten Parameter seiner Dreitupel-Bezeichnung, indem wir dessen Bezug zu dem Speziesbaum \mathcal{S} beschreiben.

5.2.1.3.1 Y'' ist ein bekanntes Gen

Ist Y'' das bekannte Gen in $E_{\mathcal{R}}(X)$, so gilt $Y'' = \varphi(X)$. Wir geben als dritten Parameter die Host-Spezies $\sqcup_{\mathcal{S}}(X)$ von Y'' an. Damit bezeichnet das Dreitupel $(X, X, \sqcup_{\mathcal{S}}(X))$ biologisch das bekannte Gen X in \mathcal{G} , dessen Host-Spezies $\sqcup_{\mathcal{S}}(X)$ ist. Die Dreitupel-Bezeichner aller in \mathcal{R} bekannten Gene sind, ohne Verwendung des Mengensystemes \mathcal{R} , durch die folgende Menge beschreibbar.

$$\{(X, X, S) \in \mathcal{G} \times \mathcal{G} \times \mathcal{S} \mid S = \sqcup_{\mathcal{S}}(X)\}$$

5.2.1.3.2 Y'' ist ein unbekanntes Gen

Ist Y'' ein unbekanntes Gen, so ist es ein Element aus einer unbestimmten oder lokalen Entwicklungsumgebung von $E_{\mathcal{R}}(X)$. Aus Abschnitt 4.4 wissen wir, daß die Topologie dieser Entwicklungsumgebungen jeweils Subtopologien von $E_{\mathcal{S}}(X)$ entsprechen. Wie sich diese Subtopologien zusammensetzen, ist davon abhängig, ob $\varphi(X)$ ein Duplikationselement ist oder nicht.

Ist $\varphi(X)$ kein Duplikationselement, so ist Y'' entweder ein Element in $e_{\mathcal{R}}(Z)$ oder $\tilde{E}_{\mathcal{R}}(X)$. Nach Korollar 4.3 gilt $B_{e_{\mathcal{R}}(Z)} \in \text{Oiso}(e_{\mathcal{R}}(Z), e_{\mathcal{S}}(Z))$ und $B_{\tilde{E}_{\mathcal{R}}(X)} \in \text{Oiso}(\tilde{E}_{\mathcal{R}}(Z), \tilde{E}_{\mathcal{S}}(X))$. Also ist Y'' eindeutig durch das Element $\sqcup_{\mathcal{S}}(Y'')$ in $e_{\mathcal{S}}(Z)$ bestimmt, wenn $Y'' \in e_{\mathcal{R}}(Z)$ gilt. Sonst ist $Y'' \in \tilde{E}_{\mathcal{R}}(X)$ und ist eindeutig durch das Element $\sqcup_{\mathcal{S}}(Y'')$ in $\tilde{E}_{\mathcal{S}}(X)$ bestimmt. Somit kennzeichnen wir Y'' eindeutig in seiner lokalen oder unbestimmten Entwicklungsumgebung durch das Dreitupel $(X, Z, \sqcup_{\mathcal{S}}(Y''))$. Die Dreitupel-Bezeichnungen aller unbekanntes Elemente der Entwicklungsumgebung $E_{\mathcal{R}}(X)$ lassen sich, ohne die Verwendung des Mengensystemes \mathcal{R} , durch die folgende Menge beschreiben.

$$\{(X, Z, S) \in \mathcal{G} \times \mathcal{G} \cup \{?\} \times \mathcal{S} \mid (X = \text{Pa}(Z) \wedge S \in e_{\mathcal{S}}(Z)) \vee (Z = ? \wedge S \in E_{\mathcal{S}}(X))\}$$

Ist $\varphi(X)$ ein Duplikationselement, so folgt $\tilde{E}_{\mathcal{R}}(X) = \emptyset$ aus Theorem 4.7. Damit ist Y'' ein Element der lokalen Entwicklungsumgebung $B_{e_{\mathcal{R}}(Z)}$. Es gilt $B_{e_{\mathcal{R}}(Z)} \in \text{Oiso}(e_{\mathcal{R}}(Z), T_S(\sqcup_S(X) \setminus \sqcup_S(Z)))$ nach Theorem 4.8. Damit besitzt $e_{\mathcal{R}}(Z)$ dieselbe Topologie wie $T_S(\sqcup_S(X) \setminus \sqcup_S(Z))$. Also ist das Element Y'' eindeutig in $e_{\mathcal{R}}(Z)$ durch $\sqcup_S(Y'')$ zu beschreiben. Somit kennzeichnen wir Y'' durch das Dreitupel $(X, Z, \sqcup_S(Y''))$. Die Dreitupel-Bezeichnungen für alle Elemente in $e_{\mathcal{R}}(Z)$, ohne die Verwendung des Mengensystemes \mathcal{R} , können wir wie folgt beschreiben.

$$\{(X, Z, S) \in \mathcal{G} \times \mathcal{G} \times \mathcal{S} \mid X = \text{Pa}(Z) \wedge S \in T_S(\sqcup_S(X) \setminus \sqcup_S(Z))\}$$

Biologisch betrachtet, beschreibt das vollständige Dreitupel (X, Z, S) mit $Z \in \text{Ch}(X)$ ein unbekanntes Gen, das durch die Entwicklung von dem bekannten Gen $(X, X, \sqcup_S(X))$ in das bekannte Gen $(Z, Z, \sqcup_S(Z))$ in der Host-Spezies S entstanden ist. Gilt $Z = ?$, so beschreibt das Dreitupel ein unbekanntes Gen, welches durch die Entwicklung von dem Gen $(X, X, \sqcup_S(X))$ zu einem Gen $(Y, Y, \sqcup_S(Y))$ mit $Y \in \text{Ch}(X)$ der Host-Spezies S entstanden ist. Welches Gen $(Y, Y, \sqcup_S(Y))$ für die Entwicklung des Gens (X, Z, S) verantwortlich war, ist somit unbestimmt.

5.2.2 Die konstruktive Reconciled Tree: formal

In Abschnitt 5.2.1 haben wir den Elementen eines Reconciled Tree \mathcal{R} Dreitupel-Bezeichner zugeordnet. Wir werden nun den konstruktiven Reconciled Tree von \mathcal{G} über \mathcal{S} als geordnete Menge von Dreitupel-Bezeichnern definieren.

Zunächst definieren wir dazu in Abschnitt 5.2.2.1 formal die Menge der Dreitupel-Bezeichner und zeigen, daß diese sich bijektiv den Elementen aus \mathcal{R} zuordnen lassen. Auf der Menge der Dreitupel-Bezeichner definieren wir dann in Abschnitt 5.2.2.2 eine Ordnung und zeigen, daß sich diese Ordnung isomorph auf \mathcal{R} überträgt.

5.2.2.1 Formale Definition der Dreitupel-Bezeichner

Die Menge der Dreitupel-Bezeichner, wie wir sie in Abschnitt 5.2.1 den Elementen aus \mathcal{R} zugeordnet haben, lassen sich formal durch die folgenden Mengen beschreiben.

Definition 5.2

Sei M eine Menge und $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$, dann definieren wir

$$\begin{aligned} \text{BG}_{\mathcal{G}, \mathcal{S}} &:= \{(X, X, S) \in \mathcal{G} \times \mathcal{G} \times \mathcal{S} \mid S = \bigsqcup_{\mathcal{S}}(X)\} \\ \text{UG}_{\mathcal{G}, \mathcal{S}} &:= \{(X, Z, S) \in \mathcal{G} \times \mathcal{G} \cup \{?\} \times \mathcal{S} \mid \\ &\quad X \in \overline{\text{DupK}} \wedge (Z \in \text{Ch}(X) \wedge S \in e_{\mathcal{S}}(Z) \\ &\quad \vee Z = ? \wedge S \in \tilde{\mathcal{E}}_{\mathcal{S}}(X)) \\ &\quad \vee X \in \text{DupK} \wedge S \in \text{T}_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \setminus \text{T}_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z))\} \\ \text{RK}_{\mathcal{G}, \mathcal{S}} &:= \text{BG}_{\mathcal{G}, \mathcal{S}} \cup \text{UG}_{\mathcal{G}, \mathcal{S}} \end{aligned}$$

Wir bezeichnen $\text{BG}_{\mathcal{G}, \mathcal{S}}$ als die bekannten Gene (abgekürzt BGs) von $\text{RK}_{\mathcal{G}, \mathcal{S}}$, $\text{UG}_{\mathcal{G}, \mathcal{S}}$ als die unbekanntenen Gene (abgekürzt UGs) von $\text{RK}_{\mathcal{G}, \mathcal{S}}$ und $\text{RK}_{\mathcal{G}, \mathcal{S}}$ als die Elemente des konstruktiven Reconciled Tree. Sind \mathcal{G}, \mathcal{S} offensichtlich, verwenden wir die Bezeichnungen BG, UG und RK.

Die Menge RK der Dreitupel-Bezeichner soll ein Äquivalent für die Elemente in \mathcal{R} darstellen. Daher zeigen wir, daß sich alle Dreitupel-Bezeichner in RK den Elementen in \mathcal{R} mit der folgenden Abbildung bijektiv zuordnen lassen.

Definition 5.3

Sei M eine Menge $\mathcal{G}, \mathcal{S} \in \mathfrak{N}$, $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$ und φ ein minimaler spezieskonformer Einbettungsmorphismus von \mathcal{G} in \mathcal{R} , dann definieren wir die Funktion

$\eta : \mathcal{R} \rightarrow \text{RK}_{\mathcal{G}, \mathcal{S}}$ mit

$$\eta(Y'') := \begin{cases} (\text{Pa}(Z), Z, B(Y'')) & \exists Z \in \mathcal{G} \setminus \{\text{Ro}(\mathcal{G})\} : Y'' \in e_{\mathcal{R}}(Z) \\ (X, ?, B(Y'')) & \exists X \in \mathcal{G} : Y'' \in \tilde{\mathcal{E}}_{\mathcal{R}}(X) \\ (X, X, \bigsqcup_{\mathcal{S}}(X)) & \exists X \in \mathcal{G} : Y'' = \varphi(X) \end{cases}$$

Lemma 5.1

Sei M eine Menge $\mathcal{G}, \mathcal{S} \in \mathfrak{N}$, $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$ und φ ein minimaler spezieskonformer Einbettungsmorphismus von \mathcal{G} in \mathcal{R} , dann ist η bijektiv.

Beweis. Nach Lemma 4.7 ist \mathcal{R} in seine Entwicklungsumgebungen partitionierbar, nach Lemma 4.11 ist eine Entwicklungsumgebung von \mathcal{R} in dessen maximales Element, dessen lokale Entwicklungsumgebungen und dessen unbestimmte Entwicklungsumgebung partitionierbar. Also gilt

$$\mathcal{R} = \varphi(\mathcal{G}) \dot{\cup} \bigcup_{X \in \mathcal{G}} (\tilde{\mathcal{E}}_{\mathcal{R}}(X)) \dot{\cup} \bigcup_{Z \in \text{Ch}(X)} e_{\mathcal{R}}(Z).$$

Für jedes Element \mathcal{P} dieser Partition zeigen wir, daß die partielle Abbildung $\eta|_{\mathcal{P}}$ bijektiv auf ihren Wertebereich abbildet. Sei $X \in \overline{\text{DupK}}$ und $Z \in \text{Ch}(X)$. Nach Korollar 4.3 sind die folgenden partiellen Abbildungen bijektiv.

$$\begin{aligned} \eta|_{e_{\mathcal{R}}(Z)} : e_{\mathcal{R}}(Z) &\rightarrow \{X\} \times \{Z\} \times e_{\mathcal{S}}(Z) \\ \eta|_{\tilde{E}_{\mathcal{R}}(X)} : \tilde{E}_{\mathcal{R}}(X) &\rightarrow \{X\} \times \{?\} \times \tilde{E}_{\mathcal{S}}(X). \end{aligned}$$

Sei $X \in \text{DupK}$ und $Z \in \text{Ch}(X)$. Nach Theorem 4.8 ist die folgende partielle Abbildung bijektiv.

$$\eta|_{e_{\mathcal{R}}(Z)} : e_{\mathcal{R}}(Z) \rightarrow \{X\} \times \{Z\} \times T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \setminus T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z)).$$

Da φ injektiv und die kleinste obere Schranke in \mathcal{S} eindeutig ist, folgt, daß die folgende partielle Abbildung bijektiv ist:

$$\eta|_{\varphi(\mathcal{G})} : \varphi(\mathcal{G}) \rightarrow \{(X, X, \mathcal{S}) \in \mathcal{G} \times \mathcal{G} \times \mathcal{S} \mid \mathcal{S} = \bigsqcup_{\mathcal{S}}(X)\}.$$

Damit wird jedes Element \mathcal{P} der Partition $\varphi(\mathcal{G}) \cup \bigcup_{Z \in \mathcal{G} \setminus \{\text{Ro}(\mathcal{G})\}} e_{\mathcal{R}}(Z)$ mit $\eta|_{\mathcal{P}}$ bijektiv in ihren Wertebereich abgebildet. Die Vereinigung der Wertebereiche dieser partitionellen Abbildungen ergeben die Menge RK.

$$\begin{aligned} \eta(\mathcal{R}) &= \eta(\varphi(\mathcal{G})) \cup \bigcup_{\substack{X \in \overline{\text{DupK}}, \\ Z \in \text{Ch}(X)}} \eta(e_{\mathcal{R}}(Z)) \cup \bigcup_{X \in \overline{\text{DupK}}} \eta(\tilde{E}_{\mathcal{R}}(X)) \\ &\cup \bigcup_{\substack{X \in \text{DupK}, \\ Z \in \text{Ch}(X)}} \eta(e_{\mathcal{R}}(Z)) \quad (\text{da } \tilde{E}_{\mathcal{R}}(X) = \emptyset \text{ für } X \in \text{DupK} \text{ (Theo. 4.7)}) \\ &= \{(X, X, \mathcal{S}) \in \mathcal{G} \times \mathcal{G} \times \mathcal{S} \mid \mathcal{S} = \bigsqcup_{\mathcal{S}}(X)\} \\ &\cup \bigcup_{\substack{X \in \overline{\text{DupK}}, \\ Z \in \text{Ch}(X)}} \{X\} \times \{Z\} \times e_{\mathcal{S}}(Z) \cup \bigcup_{X \in \overline{\text{DupK}}} \{X\} \times \{?\} \times \tilde{E}_{\mathcal{S}}(X) \\ &\cup \bigcup_{\substack{X \in \text{DupK}, \\ Z \in \text{Ch}(X)}} \{X\} \times \{Z\} \times T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X)) \setminus T_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(Z)) \\ &= \text{BG} \cup \text{UG} \\ &= \text{RK}. \end{aligned}$$

Damit ist η eine surjektive Abbildung und, da die Wertebereiche aller partiellen Abbildungen paarweise disjunkt sind, ist η injektiv. Somit ist η eine bijektive Abbildung. \square

Verw. von: Theo. 4.7, 4.8, Kor. 4.3, Lem. 4.7; Verw. in: Lem. 5.2.

5.2.2.2 Die Formale Definition einer Ordnung auf RK

In diesem Abschnitt geben wir die vollständige Definition des konstruktiven Reconciled Tree an. Dazu definieren wir auf den Dreitupel-Bezeichnern der Menge $\text{RK}_{\mathcal{G},\mathcal{S}}$ eine Ordnung, welche zu der Ordnung des Reconciled Tree \mathcal{R} ordnungsisomorph ist.

Definition 5.4 (Ordnung auf RK)

Sei M eine Menge, $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$ und $(X_i, Z_i, S_i), (X_j, Z_j, S_j) \in \text{RK}_{\mathcal{G},\mathcal{S}}$.

$$\begin{aligned} (X_i, Z_i, S_i) \subseteq (X_j, Z_j, S_j) &: \iff \\ Z_i \neq ? \quad \wedge Z_i \subseteq Z_j \wedge S_i \subseteq S_j &\quad \vee \quad (\text{aus } Z_i \subseteq Z_j \text{ folgt } Z_j \neq ?) \\ Z_i = ? \quad \wedge X_i \subseteq Z_j &\quad \vee \quad (\text{aus } X_i \subseteq Z_j \text{ folgt } Z_j \neq ?) \\ Z_i = Z_j = ? \wedge X_i = X_j \wedge S_i \subseteq S_j & \end{aligned}$$

Da $Z_j \neq ?$ aus $Z_i \subseteq Z_j$ folgt, ist die binäre Relation \subseteq auf der Menge RK wohldefiniert. Offensichtlich bildet diese binäre Relation eine Ordnung auf $\text{RK}_{\mathcal{G},\mathcal{S}}$, welche sich von \mathcal{G} und \mathcal{S} in natürlicher Weise auf $\text{RK}_{\mathcal{G},\mathcal{S}}$ überträgt. In Lemma 5.2 zeigen wir, daß die Baumordnung \subseteq auf \mathcal{R} ordnungsisomorph zu der binären Relation \subseteq auf der Menge $\text{RK}_{\mathcal{G},\mathcal{S}}$ ist.

Definition 5.5 (Konstruktiver Reconciled Tree)

Sei M eine Menge, $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$, dann bezeichnen wir $\text{RK}_{\mathcal{G},\mathcal{S}}$ mit der Ordnung \subseteq als den konstruktiven Reconciled Tree von \mathcal{G} über \mathcal{S} .

Lemma 5.2

Sei $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$, dann gilt $\eta \in \text{Oiso}(\mathcal{R}, \text{RK}_{\mathcal{G},\mathcal{S}})$.

Beweis. Nach Lemma 5.1 ist η bijektiv, womit $\eta \in \text{Ohom}(\mathcal{R}, \text{RK})$ zu zeigen bleibt. Seien dazu $Y_i'', Y_j'' \in \mathcal{R}$ mit $Y_i'' \subset Y_j''$, $\eta(Y_i'') = (X_i, Z_i, S_i)$ und $\eta(Y_j'') = (X_j, Z_j, S_j)$. Damit ist $Y_i'' \in \mathbf{E}_{\mathcal{R}}(X_i)$ und $Y_j'' \in \mathbf{E}_{\mathcal{R}}(X_j)$. Wir werden $\eta(Y_i'') \subseteq \eta(Y_j'')$ zeigen. Dazu machen wir die beiden folgenden Aussagen.

Aussage 1: Wir zeigen, daß allgemein $S_i \subseteq S_j$ gilt. Dazu zeigen wir zuerst, daß $S_i = \mathbf{B}(Y_i'')$ und $S_j = \mathbf{B}(Y_j'')$ gilt. Dazu sei $\eta(Y'') = (X, Z, S)$ für ein festes, aber beliebiges $Y'' \in \mathcal{G}$. Falls $Y'' = \varphi(X)$ gilt, ist $S = \bigsqcup_{\mathcal{S}}(X)$, sonst ist $S = \mathbf{B}(Y'')$. Gilt $Y'' = \varphi(X)$, so ist $\mathbf{B}(Y'') = \mathbf{B}(\varphi(X))$. Aus dem LCA-Theorem 4.4 folgt, daß $\mathbf{B}(\varphi(X)) = \bigsqcup_{\mathcal{S}}(X)$ und damit $S = \mathbf{B}(Y'')$ folgt. Also gilt in jedem Fall $S = \mathbf{B}(Y'')$.

Im besonderen gilt somit $S_i = \mathbf{B}(Y_i'')$ und $S_j = \mathbf{B}(Y_j'')$. Aus Lemma 3.1 folgt $\mathbf{B}(Y_i'') \subseteq \mathbf{B}(Y_j'')$ aus $Y_i'' \subseteq Y_j''$, womit $S_i \subseteq S_j$ gezeigt ist.

Aussage 2: Aus $Y_i'' \in \tilde{E}_{\mathcal{R}}(X_i)$ folgt $Z_i = ?$, aus $Y_i'' = \bigsqcup_S(X_i)$ folgt $Z_i = X_i$ und sonst ist $Y'' \in e_{\mathcal{R}}(Z)$ mit $Z \in \text{Ch}(X_i)$, woraus $Z_i = Z$ folgt. Adäquates gilt für Y_j'' .

Wir untersuchen nun die folgenden vier Fälle für Y_i'' und Y_j'' : (i) $Y_i'' \in \tilde{E}_{\mathcal{R}}(X_i)$, $Y_j'' \in \tilde{E}_{\mathcal{R}}(X_j)$, (ii) $Y_i'' \notin \tilde{E}_{\mathcal{R}}(X_i)$, $Y_j'' \in \tilde{E}_{\mathcal{R}}(X_j)$, (iii) $Y_i'' \in \tilde{E}_{\mathcal{R}}(X_i)$, $Y_j'' \notin \tilde{E}_{\mathcal{R}}(X_j)$, (iv) $Y_i'' \notin \tilde{E}_{\mathcal{R}}(X_i)$, $Y_j'' \notin \tilde{E}_{\mathcal{R}}(X_j)$.

Zu den Fällen (i) und (ii): Alle kleinsten Elemente in $\tilde{E}_{\mathcal{R}}(X_j)$ sind Blätter. Aus $Y_j'' \in \tilde{E}_{\mathcal{R}}(X_j)$ und $Y_i'' \subseteq Y_j''$ folgt somit $X_i = X_j$. Damit existiert Fall (ii) nicht. Gilt Fall (ii), so folgt aus Aussage 1 und Aussage 2, daß $S_i \subseteq S_j$ und $Z_i = Z_j = ?$ gilt. Da $X_i = X_j$ gilt, folgt $\eta(Y_i'') \subseteq \eta(Y_j'')$.

Zu den Fällen (iii) und (iv): Gilt Fall (iii), so gilt $Z_i = ?$ und $Z_j \neq ?$ nach Aussage 2. Wenn $X_i \subseteq Z_j$ gilt, folgt $\eta(Y_i'') \subseteq \eta(Y_j'')$.

Gilt Fall (iv), so gilt $Z_i \neq ?$ nach Aussage 2, nach Aussage 1 gilt $S_i \subseteq S_j$. Wenn $Z_i \subseteq Z_j$ gilt, folgt $\eta(Y_i'') \subseteq \eta(Y_j'')$.

Es verbleibt somit, $X_i \subseteq Z_j$ für Fall (iii) und $S_i \subseteq S_j$ für Fall (iv) zu zeigen.

Da die Entwicklungsumgebungen \mathcal{R} partitionieren (Lem. 4.7), gilt entweder $E_{\mathcal{R}}(X_i) = E_{\mathcal{R}}(X_j)$ oder $E_{\mathcal{R}}(X_i) \cap E_{\mathcal{R}}(X_j) = \emptyset$. Diese beiden Fälle untersuchen wir nun im folgenden auf die zu zeigenden Eigenschaften.

1. $E_{\mathcal{R}}(X_i) \cap E_{\mathcal{R}}(X_j) = \emptyset$:

Aussage 3: Wir zeigen $\varphi(X_i) \subset Y_j''$. Es gilt $Y_i'' \subset Y_j''$, und aus $Y_i'' \in E_{\mathcal{R}}(X_i)$ folgt $Y_i'' \subseteq \varphi(X_i)$. Damit sind Y_j'' und $\varphi(X_i)$ durch \subseteq vergleichbar, womit $Y_j'' \subseteq \varphi(X_i)$ oder $\varphi(X_i) \subset Y_j''$ gilt. Nehmen wir an, es gilt $Y_j'' \subseteq \varphi(X_i)$. Mit $Y_i'' \subset Y_j''$ und $Y_i'' \subseteq \varphi(X_i)$ folgt $Y_j'' \in K_{\mathcal{R}}(\varphi(X_i), Y_i'')$. Da $\varphi(X_i), Y_i'' \in E_{\mathcal{R}}(X_i)$ sind, ist $K_{\mathcal{R}}(\varphi(X_i), Y_i'') \subseteq E_{\mathcal{R}}(X_i)$. Damit ist aber $Y_j'' \in E_{\mathcal{R}}(X_i)$, womit $E_{\mathcal{R}}(X_i) \cap E_{\mathcal{R}}(X_j) \neq \emptyset$ im Widerspruch zu $E_{\mathcal{R}}(X_i) \cap E_{\mathcal{R}}(X_j) = \emptyset$ folgt. Somit gilt $\varphi(X_i) \subset Y_j''$.

Aussage 4: Wir zeigen, daß $\varphi(Z_i) \subseteq \varphi(X_i)$ für $Y_i'' \notin \tilde{E}_{\mathcal{R}}(X_i)$ gilt. Wenn $Y_i'' = \varphi(X_i)$ ist, folgt $Z_i = X_i$ und somit $\varphi(Z_i) = Y_j''$. Sonst ist $Y_i'' \in e_{\mathcal{R}}(Z_i)$, da $Y_i'' \notin \tilde{E}_{\mathcal{R}}(X_i)$, woraus $\varphi(Z_i) \subset \varphi(X_i)$ folgt. Also gilt $\varphi(Z_i) \subseteq \varphi(X_i)$.

Wir unterscheiden nun die beiden Fälle $\varphi(X_j) = Y_j''$ und $\varphi(X_j) \neq Y_j''$, welche in Abbildung 5.5 für den Fall $Y_i'' \notin \tilde{E}_{\mathcal{R}}(X_i)$ und $Y_j'' \notin \tilde{E}_{\mathcal{R}}(X_j)$

veranschaulicht sind. Hierbei ist $\varphi(I)$ für ein Element $I \in \mathcal{G}$ durch I'' gekennzeichnet (z.B. $\varphi(Z_i) = Z_i''$).

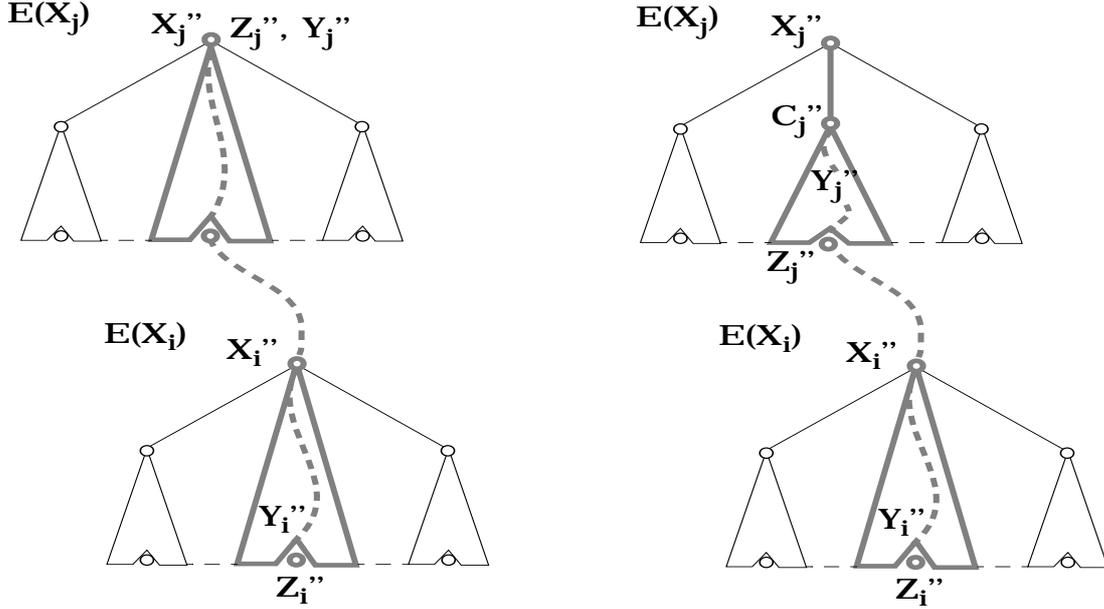


Abbildung 5.5: Fall 1a links und Fall 1b rechts

(a) $\varphi(X_j) = Y_j''$:

Fall (iv): Damit folgt $X_j = Z_j$ nach Aussage 2 und weiter $Y_j'' = \varphi(Z_j)$. Aus $\varphi(Z_i) \subseteq \varphi(X_i)$ (Aussage 3), $\varphi(X_i) \subset Y_j''$ (Aussage 4) und $Y_j'' = \varphi(Z_j)$ folgt $\varphi(Z_i) \subset \varphi(Z_j)$. Da φ einbettungsmonomorph ist, folgt hieraus $Z_i \subset Z_j$.

Fall (iii): Wie im Fall (iv) gilt $\varphi(X_i) \subset Y_j''$ (Aussage 4) und $Y_j'' = \varphi(Z_j)$, woraus $\varphi(X_i) \subset \varphi(Z_j)$ folgt. Da φ einbettungsmonomorph ist, folgt hieraus $X_i \subset Z_j$.

(b) $\varphi(X_j) \neq Y_j''$:

Fall (iv): Damit ist $Y_j'' \in e_{\mathcal{R}}(Z_j)$. $\varphi(Z_j)$ ist das einzige Kind eines Elementes in $e_{\mathcal{R}}(Z_j)$, welches nicht in $e_{\mathcal{R}}(Z_j)$ enthalten ist. Also ist jedes Element in \mathcal{R} , das nicht in $e_{\mathcal{R}}(Z_j)$ enthalten, aber Teilmenge eines Elementes in $e_{\mathcal{R}}(Z_j)$ ist, Teilmenge von $\varphi(Z_j)$.

Es gilt $E_{\mathcal{R}}(X_i) \cap E_{\mathcal{R}}(X_j) = \emptyset$, woraus $\varphi(X_i) \notin E_{\mathcal{R}}(X_j)$ und damit $\varphi(X_i) \notin e_{\mathcal{R}}(Z_j)$ folgt. Mit $\varphi(X_i) \subset Y_j''$ ist dann $\varphi(X_i)$ nicht in $e_{\mathcal{R}}(Z_j)$ enthalten, aber Teilmenge eines Elementes in $e_{\mathcal{R}}(Z_j)$. Hieraus folgt nun $\varphi(X_i) \subseteq \varphi(Z_j)$. Mit $\varphi(Z_i) \subseteq \varphi(X_i)$ folgt dann $\varphi(Z_i) \subseteq \varphi(Z_j)$. Da φ einbettungsmonomorph ist, folgt daraus $Z_i \subseteq Z_j$.

Fall (iii): Wie im Fall (iV) folgt $\varphi(X_i) \subseteq \varphi(Z_j)$. Da φ einbettungsmonomorph ist, folgt daraus $X_i \subseteq Z_j$.

Somit gilt $Z_i \subseteq Z_j$ im Fall (iV) und $X_i \subseteq Z_j$ im Fall (iii)..

2. $E_{\mathcal{R}}(X_i) = E_{\mathcal{R}}(X_j)$: Damit ist $X_i = X_j$.

Fall (iii): Damit ist $X_i = X_j$. Es folgt $Y_i'' \in \tilde{E}_{\mathcal{R}}(X_i)$ und $Y_i'' \notin \tilde{E}_{\mathcal{R}}(X_i)$. Mit $Y_i'' \subseteq Y_j''$, folgt hieraus $Y_j'' = \bigsqcup_{\mathcal{S}}(X_i)$. Nach Aussage 2 ist damit $X_i = Z_j$.

Fall (iv): Es gilt $X_i = X_j$, womit $Y_i'', Y_j'' \in E_{\mathcal{R}}(X_j)$ folgt. Aus $Y_i'' \subset Y_j''$ folgt $Y_i'' \neq \varphi(X_j)$. Also ist $Y_i'' \in e_{\mathcal{R}}(Z_i)$ und $\text{Pa}(Z_i) = X_j$. Wir unterscheiden nun die beiden Fälle $Y_j'' = \varphi(X_j'')$ und $Y_j'' \neq \varphi(X_j'')$, welche in Abbildung 5.6 noch einmal veranschaulicht sind (wieder ist dabei die Abbildung eines Elementes $I \in \mathcal{G}$ unter φ durch I'' gekennzeichnet).

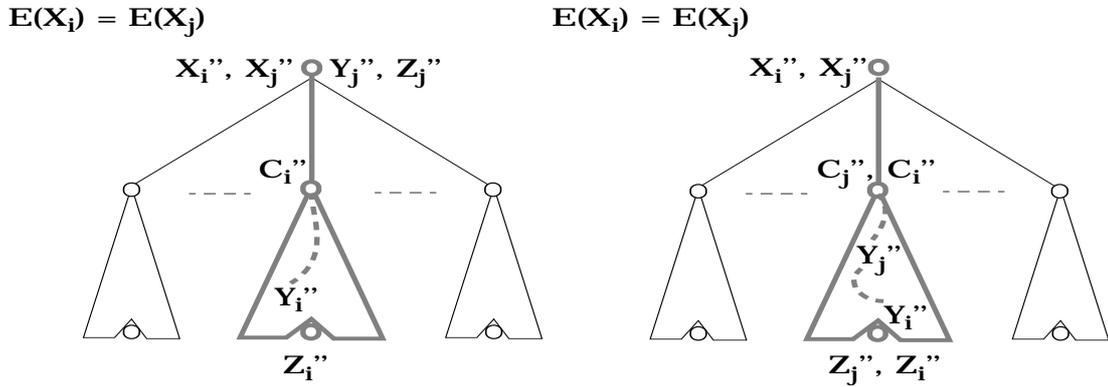


Abbildung 5.6: Fall 2a links und Fall 2b rechts

- (a) $Y_j'' = \varphi(X_j)$: Damit ist $Z_j = X_j$, und mit $\text{Pa}(Z_i) = X_j$ folgt $Z_i \subset Z_j$.
 (b) $Y_j'' \neq \varphi(X_j)$: Damit ist $Y_j'' \in e_{\mathcal{R}}(Z_j)$ und $\text{Pa}(Z_j) = X_j$. Es gilt $Y_i'' \in e_{\mathcal{R}}(Z_i)$ und $\text{Pa}(Z_i) = X_j$. Die Elemente unterschiedlicher lokaler Entwicklungsumgebungen einer Entwicklungsumgebung in \mathcal{R} sind nicht vergleichbar. Y_i'' und Y_j'' sind jedoch durch \subset vergleichbar, womit sie in derselben lokalen Entwicklungsumgebung enthalten sein müssen. Damit gilt $Z_i = Z_j$.

Somit gilt $X_i \subseteq Z_j$ im Fall (iii) und $Z_i \subseteq Z_j$ im Fall (iv).

Damit ist $\eta \in \text{Oiso}(\mathcal{R}, \text{RK}(\mathcal{G}, \mathcal{S}))$. □

Verw. von: Theo. 4.4, Lem. 3.1, 4.7, 5.1; Verw. in: .

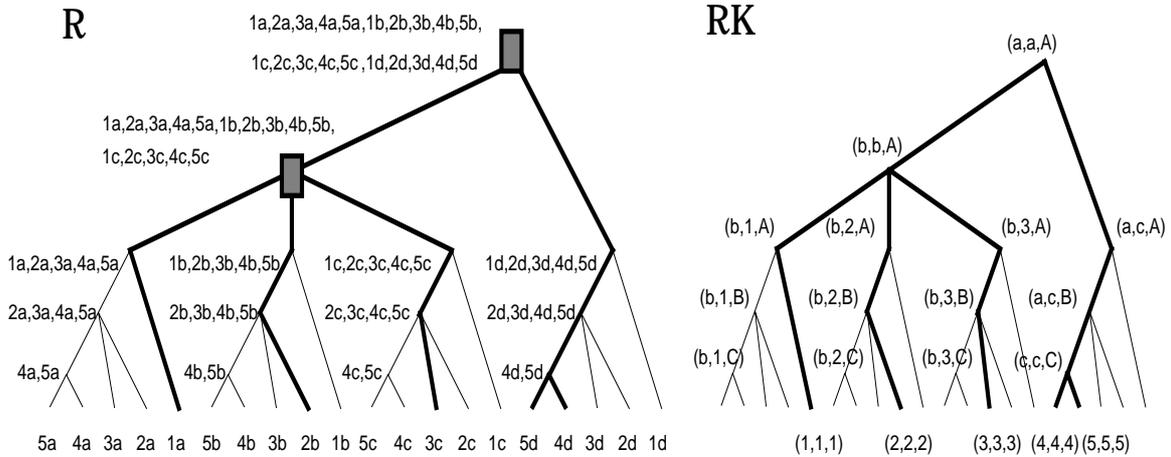


Abbildung 5.7: Ein Reconciled Tree und der konstruktive Reconciled Tree

Der konstruktive Reconciled Tree ist anschaulicher und auch formal einfacher zu handhaben als ein nicht konstruktiv definierter Reconciled Tree. Ein Vergleich zeigt Abbildung 5.7.

Daher werden wir von nun an im wesentlichen den konstruktiv definierten Reconciled Tree verwenden. Die folgenden Eigenschaften des Reconciled Tree übertragen sich auf den konstruktiven Reconciled Tree.

Wir übertragen die Basisoperation B von der Multimenge auf die Dreitupel von $RK(\mathcal{G}, \mathcal{S})$.

Definition 5.6 (Basis einer Dreitupel-Bezeichnung)

Sei M eine Menge und $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$ und $(X, Z, S) \in RK_{\mathcal{G}, \mathcal{S}}$, dann bezeichnen wir $B((X, Z, S)) := S$ als die Basis des Dreitupels (X, Z, S) .

Mit dieser auf Dreitupeln definierten Basisfunktion übertragen sich die Definitionen des Duplikationselementes der Ableitbarkeit und der spezieskonformen Einbettung des nicht konstruktiv definierten Reconciled Tree auf den konstruktiv definierten Reconciled Tree. Aus dem Reconciled Tree leitet sich auch die Einbettung des Genbaumes \mathcal{G} in RK ab.

Lemma 5.3

Sei M eine Menge und $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$, dann existiert ein $\varphi \in \text{Eiso}(\mathcal{G}, RK_{\mathcal{G}, \mathcal{S}})$.

Beweis. Es gilt $RC(\mathcal{G}, \mathcal{S})$ ist nicht leer, womit ein $\mathcal{R} \in RC(\mathcal{G}, \mathcal{S})$ existiert. Nach Definition von \mathcal{R} existiert eine minimale Abbildung $\varphi \in \text{Emon}(\mathcal{G}, \mathcal{R})$. Lemma 5.2 besagt, daß ein $\eta \in \text{Oiso}(\mathcal{R}, RK(\mathcal{G}, \mathcal{S}))$ existiert. Somit ist $\eta \circ \varphi \in \text{Emon}(\mathcal{G}, RK(\mathcal{G}, \mathcal{S}))$. \square

Verw. von: Lem. 5.2; Verw. in: .

Mit einem minimalen Einbettungsmonomorphismus $\varphi \in \text{Eiso}(\mathcal{G}, \text{RK}_{\mathcal{G}, \mathcal{S}})$ übertragen sich die Definitionen der Entwicklungsumgebung und der lokalen Entwicklungsumgebung von dem Reconciled Tree auf den konstruktiven Reconciled Tree.

Damit kann der konstruktiv definierte Reconciled Tree im wesentlichen wie der nicht konstruktiv definierte Reconciled Tree verwendet werden. Im weiteren Sprachgebrauch verstehen wir unter dem Reconciled Tree den konstruktiv definierten Reconciled Tree.

5.3 Ausblick: Der Mixed Tree

Die konstruktive Definition der Reconciled Tree RK ist eine biologisch einfach interpretierbare Rekonstruktion der Gene des Genbaumes \mathcal{G} unter der Berücksichtigung aller postulierter Genduplikationen, welche die Inkonsistenz zwischen \mathcal{G} und \mathcal{S} erklären. Die Praxis zeigt, daß schon bei wenigen postulierten Genduplikationen der Reconciled Tree unübersichtlich werden kann (siehe z.B. Abb. 8.7 in Kapitel 8). Oft interessiert den Biologen nur lokal die Auswirkung von postulierten Genduplikationen, und er möchte aufgrund der so gewonnenen Erkenntnisse den Genbaum verändern.

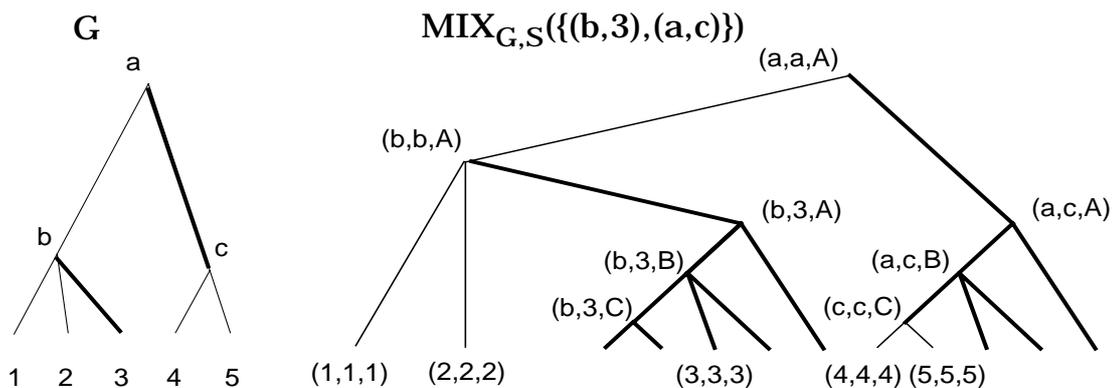


Abbildung 5.8: Ein Mixed Tree

Wir haben gezeigt, daß Kanten des Genbaumes \mathcal{G} mögliche unbekannte Gen-Entwicklungen *verbergen*, indem diese durch nicht erkannte Genduplikationen im Genbaum unentdeckt bleiben. Der Reconciled Tree berücksichtigt die Genduplikationen und zeigt dadurch die durch Kanten des Genbaumes verborgenen Gen-Entwicklungen. Befindet sich diese Entwicklung in einer lokalen Entwicklungsumgebung $e_{\mathcal{R}}(Z)$, so verbirgt die Kante $(\text{Pa}(Z), Z)$ diese Entwicklung in

\mathcal{G} . Sonst befindet sich die unbekannte Entwicklung in einer unbestimmten Entwicklungsumgebung $\tilde{E}_{\mathcal{R}}(X)$, und wir wissen nur, daß diese Entwicklung durch Kanten von X zu den Kindern $\text{Ch}(X)$ in \mathcal{G} verborgen sind. Jedoch wissen wir nicht, welche dieser Kanten dafür verantwortlich sind.

Damit sind Mischformen zwischen dem Reconciled Tree RK und dem Genbaum \mathcal{G} möglich, indem wir einen Teil der durch Kanten verborgenen Gen-Entwicklungen in \mathcal{G} aufzeigen. Solche Mischformen bezeichnen wir als *Mixed Trees*. Die Abbildung 5.8 zeigt einen Mixed Tree, bei dem die verborgenen Gen-Entwicklungen der Kanten (a, b) und $(b, 3)$ aufgezeigt sind. Die verborgenen Gen-Entwicklungen der verbleibenden Kanten ist nicht aufgezeigt.

Die Knoten eines Mixed Tree werden durch die Dreitupel-Bezeichnung des Reconciled Tree gekennzeichnet. Stellt dieser den Genbaum \mathcal{G} dar, so sind alle Gene bekannt, und wir bezeichnen die Knoten des Mixed Tree mit den Dreitupel-Bezeichnern der bekannten Gene BG. Besitzt ein Mixed Tree aufgelöste Kanten, so werden diese durch die Dreitupel-Bezeichner aus UG der zugehörigen lokalen oder unbestimmten Entwicklungsumgebungen von RK bezeichnet. Der unaufgelöste Mixed Tree beschreibt somit \mathcal{G} , und der vollkommen aufgelöste Mixed Tree den Reconciled Tree RK.

Der Mixed Tree ist ein Resultat der konstruktiven Beschreibung des Reconciled Tree. Ideen zu seiner direkten Berechnung sowie zu Maßen, welche dessen Qualität beschreiben, existieren zur Zeit nur in Ansätzen. Vorstellbar ist hier, daß der Mixed Tree nicht anhand des vorher erstellten Reconciled Tree dargestellt wird, sondern dieser für die konkrete Auflösung seiner Kanten explizit berechnet wird. Dies hat den Vorteil, daß der Biologe den Mixed Tree als "Modelling Tool" zur Rekonstruktion des Genbaumes unter der Berücksichtigung von Gen-duplikationen einsetzen kann. Unterstützt durch Maße zum Mixed Tree kann der Biologe nach seinen Erfahrungen den Genbaum rekonstruieren, indem er die Auswirkungen lokaler Veränderungen direkt analysiert. Das "starre" Modell des Reconciled Tree wird somit durch den Mixed Tree dynamisiert.

Kapitel 6

Die biologischen Maße des Reconciled Tree

In der Anwendung sind wir an einem Maß interessiert, welches die Inkonsistenz von einem Gen- und Speziesbaum $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$ für eine Menge M beschreibt. Hierzu könnten wir eines der üblichen phylogenetischen Maße, wie Partition Metrik (Borque, 1978; Robinson und Foulds, 1981), Quartet-Metrik (Estabrook, 1985) oder Nearest-Neighbour-Interchange-Metrik (Robinson, 1971; Smith und Waterman, 1978) verwenden. Diese Maße vergleichen Bäume jedoch auf einer mathematisch abstrakten Ebene, in welche wir erst eine biologische Erklärung hinein interpretieren müssen. Der Reconciled Tree $RK_{\mathcal{G}, \mathcal{S}}$ gibt jedoch schon eine biologische Erklärung für die Inkonsistenz von \mathcal{G} und \mathcal{S} , aus welcher wir die Maße direkt ableiten werden.

Dazu beginnen wir mit der Definition eines "groben" Maßes, um dieses dann schrittweise zu verfeinern. Allgemein gibt es zwei Konzepte, um Maße zum Reconciled Tree zu definieren. Das erste Konzept definiert die Maße direkt und das zweite indirekt auf dem Reconciled Tree. Indirekt wird ein Maß auf dem Gen- und Speziesbaum definiert, durch welche der Reconciled Tree bestimmt wird. Für die Berechnung eines Maßes wird meist dessen indirekte Definition verwendet, da hierfür nicht erst der Reconciled Tree explizit erstellt werden muß. Die direkte Definition eines Maßes erklärt dagegen dessen biologische Bedeutung oft besser als dessen indirekte Definition. In Abschnitt 6.1 werden wir daher die Maße zum Reconciled Tree zuerst direkt definieren und dann jeweils deren äquivalente indirekte Darstellung zeigen.

1979 ist von dem Systematiker Goodman et al. das erste Maß zum Reconciled Tree eingeführt worden. Bis heute sind eine Reihe weiterer Maße, oft unabhängig voneinander, entstanden. Dies hat dazu geführt, daß Maße sel-

ten schrittweise in ihrer chronologischen Entstehung verfeinert wurden. Einige Autoren haben ihre Maße nur nach einem Konzept, meist in Unkenntnis des anderen Konzeptes¹, definiert. Dadurch sind gleiche Definitionen jeweils indirekt und direkt als unterschiedliches Maß in die Literatur eingeführt worden. In Abschnitt 6.2 werden wir die chronologische Entwicklung der Maße zum Reconciled Tree zusammenfassend beschreiben. Dabei werden wir die Äquivalenz zwischen wichtigen historischen und in Abschnitt 6.1 definierten Maßen zeigen. Einerseits erfährt eine oft informale historische Definition hierdurch eine formale Interpretation. Andererseits ergibt sich die äquivalente Darstellung des historischen Maßes nach dem jeweils anderen Definitions-Konzept. Dadurch werden dann äquivalente Definitionen aufgedeckt. Damit zeigt sich, daß den hier betrachteten historischen Maßen ein gleiches biologischen Konzept zugrunde liegt. Die unterschiedlichen Definitionen sind im wesentlichen nur unterschiedliche Verfeinerungen dieses biologischen Konzepts.

1995 führten Mirkin et al. (1995) das Maß M_{MMS} ein, welches sich konzeptionell von den bisherigen Maßen deutlich unterschied. Umso erstaunlicher war es daher, daß in der Anwendung das Maß M_{MMS} und ein nach dem bisherigen Konzept definiertes Maß immer dieselben Werte ergaben. Hieraus leiteten Mirkin et al. (1995) die Vermutung ab, daß diese beiden Maße äquivalent sind. Diese unter dem Namen Mirkin-Muchnik-Smith-Vermutung bekannte Vermutung ist von Eulenstein und Vingron (1995) das erste Mal gezeigt worden. Später ist hierzu ein unterschiedlicher Beweis von Zhang (1997) gezeigt worden. Varianten des ursprünglichen Beweises von Eulenstein und Vingron sind in Eulenstein, Mirkin und Vingron (1997) sowie Eulenstein und Vingron (1998) zu finden. In Abschnitt 6.3 zeigen wir die Idee des Beweises von Eulenstein und Vingron.

Daß die Entwicklung der Maße zum Reconciled Tree noch lange nicht abgeschlossen ist, werden wir in Abschnitt 6.4 zeigen, indem wir neue Ansätze für zukünftige Entwicklungen von Maßen angeben.

6.1 Maße für UGs

Der Reconciled Tree $RK_{G,S}$ erklärt die Inkonsistenz zwischen dem Genbaum G und dem Speziesbaum S , indem dieser die Phylogenie der Gene aus G unter der Berücksichtigung von Genduplikationen darstellt. Sind G und S inkonsistent,

¹Dies liegt teilweise daran, daß der Reconciled Tree zum Zeitpunkt der Definition des Maßes noch nicht explizit definiert gewesen ist.

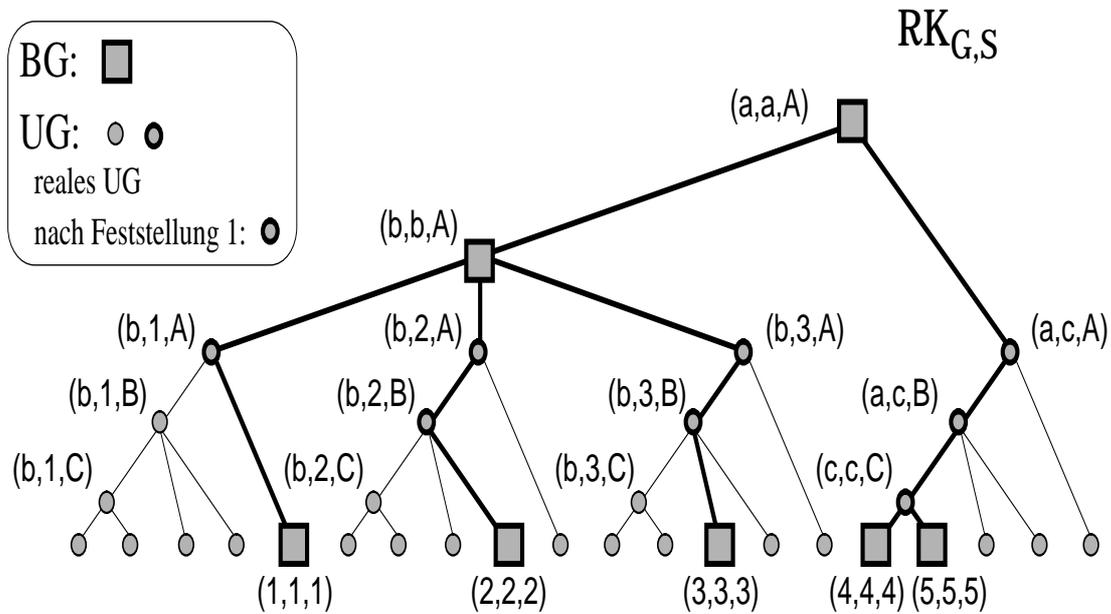


Abbildung 6.1: BGs und UGs in einem konstruktiven Reconciled Tree

so haben wir mit Genduplikationen die Ableitung RK aus \mathcal{S} erzeugt, in welche \mathcal{G} spezieskonform einbettbar ist. Die Einbettbarkeit von \mathcal{G} in RK haben wir jedoch durch zusätzliche aus \mathcal{S} abgeleitete Gene (UGs) “erkauft”, in welche kein Gen aus \mathcal{G} einbettbar ist. Damit beschreiben die unbekanntenen Gene in RK die Inkonsistenz zwischen \mathcal{G} und \mathcal{S} . Ein Beispiel gibt Abbildung 6.1. Durch die erste Duplikation an dem Duplikationselement (a, a, A) wird der Teilgraph $\mathcal{G} \setminus T_{\mathcal{G}}(b)$ in RK einbettbar. Diese Einbettung wird durch die UGs in dem Teilbaum $T_{RK}((a, c, A))$ “erkauft”. Der verbleibende Teilbaum $T_{\mathcal{G}}(b)$ wird durch die Duplikation an dem Duplikationselement (b, b, A) in RK einbettbar, welches durch die UGs in dem Teilbaum $T_{RK}((b, b, A))$ “erkauft” wird. Durch die UGs ist \mathcal{G} in die Ableitung RK aus \mathcal{S} einbettbar und damit seine Inkonsistenz zu \mathcal{S} durch Genduplikationen erklärt worden.

Wir können die Inkonsistenz zwischen \mathcal{G} und \mathcal{S} auch durch die Inkonsistenz zwischen dem Genbaum \mathcal{G} und seiner Darstellung durch RK mit Genduplikationen beschreiben. Die Inkonsistenz zwischen \mathcal{G} und RK wird durch UGs beschrieben, und die UGs beschreiben die Inkonsistenz zwischen \mathcal{G} und \mathcal{S} . Damit ist die Inkonsistenz zwischen \mathcal{G} und \mathcal{S} durch die Inkonsistenz zwischen \mathcal{G} und \mathcal{R} zu beschreiben.

Die wesentliche Beschreibung der Inkonsistenz erfolgt jedoch durch die UGs, mit welchen wir die Maße für den Reconciled Tree im folgenden definieren. In Abschnitt 6.1.1 beschreiben wir mit den UGs ein grobes Maß, welches dann in Abschnitt 6.1.2 verfeinert wird, indem wir die UGs genauer analysieren.

6.1.1 Ein einfaches Maß für UGs

Das sehr grobe Maß M_0 drückt den Unterschied zwischen \mathcal{G} und \mathcal{R} einfach durch die Anzahl der UGs, $|\text{UG}|$, aus. Aus $\text{RK} = \text{BG} \dot{\cup} \text{UG}$ und $\text{BG} = \varphi(\mathcal{G})$ ergibt sich $|\text{UG}| = |\text{RK}| - |\mathcal{G}|$. Damit definieren wir das Maß M_0 .

Definition 6.1 (M_0)

Sei M eine Menge, $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$ und $\text{RK}_{\mathcal{G}, \mathcal{S}}$ der Reconciled Tree von \mathcal{G} über \mathcal{S} . Wir bezeichnen

$$M_0(\mathcal{G}, \mathcal{S}) := |\text{RK}_{\mathcal{G}, \mathcal{S}}| - |\mathcal{G}|$$

als das 0-Maß.

RK beinhaltet alle formal möglichen UGs. Damit beschreibt RK die Entwicklung von UGs, welche wirklich einmal existiert haben, aber auch die theoretisch mögliche Weiterentwicklung von ausgestorbenen UGs. UGs, welche wirklich einmal existiert haben, bezeichnen wir als *reale UGs*. Alle anderen UGs haben niemals existiert, womit wir diese als *irreale UGs* bezeichnen. Biologisch macht es keinen Sinn, die Inkonsistenz zwischen \mathcal{G} und \mathcal{S} durch irreale UGs auszudrücken. Daher werden wir versuchen uns bei der weiteren Formulierung von Maßen nur auf die realen UGs zu beschränken. Das 0-Maß gibt hierzu nur eine obere Schranke an. In dem nächsten Abschnitt werden wir Maße einführen, welche die realen UGs besonders berücksichtigen.

6.1.2 Maße realer UGs

Wir werden nun Maße formulieren, welche möglichst nur die realen UGs berücksichtigen. Hierzu benötigen wir Aussagen darüber, wann ein UG real und wann unreal ist. Dies formulieren wir durch die beiden offensichtlichen Feststellungen:

1. Ein UG ist real, wenn es sich in RK zu einem BG entwickelt hat.
2. Ein UG ist unreal, wenn es sich in RK aus einem ausgestorbenen Gen entwickelt hat.

Aus diesen beiden Feststellungen leiten wir im folgenden die Maße M_1 , M_f und $M_{1,f}$ ab.

6.1.2.1 Das Maß M_1

Für das Maß M_1 setzen wir voraus, daß reale UGs nur aufgrund von Feststellung 1 existieren. Ein UG ist somit genau dann real, wenn sich dieses zu einem

BG entwickelt hat. Damit liegen alle realen UGs auf Pfaden zwischen BGs und sind somit Elemente reduzierter Ketten $K_{\text{RK}}^-(\varphi(X), \varphi(Z))$ für $X \in \mathcal{G}$ und $Z \in \text{Ch}(X)$. Ein Beispiel: In Abbildung 6.1 beschreiben die reduzierten Ketten

$$\begin{aligned} K_{\text{RK}}^-((b, b, A), (1, 1, 1)) &= \{(b, 1, A)\}, \\ K_{\text{RK}}^-((b, b, A), (2, 2, 2)) &= \{(b, 2, A), (b, 2, B)\}, \\ K_{\text{RK}}^-((b, b, A), (3, 3, 3)) &= \{(b, 3, A), (b, 3, B)\}, \\ K_{\text{RK}}^-((a, a, A), (c, c, C)) &= \{(a, c, A), (a, c, B)\}, \end{aligned}$$

nach Feststellung 1, alle realen UGs. Das Maß M_1 gibt den Unterschied zwischen \mathcal{G} und RK durch die Anzahl $|\bigcup_{X \in \mathcal{G}, Z \in \text{Ch}(X)} K_{\text{RK}}^-(\varphi(X), \varphi(Z))|$ der UGs, nach Feststellung 1, an. Da

$$K_{\text{RK}}^-(\varphi(\text{Pa}(Z_i)), \varphi(Z_i)) \cap K_{\text{RK}}^-(\varphi(\text{Pa}(Z_j)), \varphi(Z_j)) = \emptyset$$

für ungleiche $Z_i, Z_j \in \mathcal{G} \setminus \{\text{Ro}(\mathcal{G})\}$ gilt, definieren wir M_1 wie folgt.

Definition 6.2 (M_1)

Sei M eine Menge, $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$ und \mathcal{R} der Reconciled Tree von \mathcal{G} über \mathcal{S} . Wir definieren das 1-Maß wie folgt.

$$M_1(\mathcal{G}, \mathcal{S}) := \sum_{X \in \mathcal{G}, Z \in \text{Ch}(X)} |K_{\text{RK}}^-(\varphi(X), \varphi(Z))|$$

Da sich RK aus den Topologien von \mathcal{G} und \mathcal{S} zusammensetzt, läßt sich M_1 ohne RK (indirekt) berechnen.

Lemma 6.1

Sei M eine Menge $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$, $X \in \mathcal{G}$, $Z \in \text{Ch}(X)$ und

$$l(X, Z) := \begin{cases} |K_{\mathcal{S}}^-(\sqcup_{\mathcal{S}}(X), \sqcup_{\mathcal{S}}(Z))| + 1 & X \in \text{DupK} \wedge \sqcup_{\mathcal{S}}(Z) \neq \sqcup_{\mathcal{S}}(X) \\ |K_{\mathcal{S}}^-(\sqcup_{\mathcal{S}}(X), \sqcup_{\mathcal{S}}(Z))| & X \notin \text{DupK} \\ 0 & \text{sonst.} \end{cases}$$

Dann ist

$$M_1(\mathcal{G}, \mathcal{S}) = \sum_{X \in \mathcal{G}, Z \in \text{Ch}(X)} l(X, Z).$$

Beweis. Wir beweisen $M_1 = \sum_{X \in \mathcal{G}, Z \in \text{Ch}(X)} l(X, Z)$, indem wir für ein festes aber beliebiges $X \in \mathcal{G}$ und $Z \in \text{Ch}(X)$ die folgende Aussage zeigen:

$$l(X, Z) = |K_{\text{RK}}^-(\varphi(X), \varphi(Z))|.$$

Hierzu untersuchen wir die drei für 1 möglichen Fälle.

1. Gilt $\sqcup_s(Z) = \sqcup_s(X)$ (und $X \in \text{DupK}$)², so folgt

$$K_{\text{RK}}^-(\varphi(X), \varphi(Z)) = K_{\text{RK}}^-((X, X, \sqcup_s(X)), (X, X, \sqcup_s(X))) = \emptyset,$$

womit $|K_{\text{RK}}^-(\varphi(X), \varphi(Z))| = 0 = 1(X, Z)$ folgt.

2. Gilt $\sqcup_s(Z) \neq \sqcup_s(X)$ und $X \in \text{DupK}$, so ist

$$e_{\text{RK}}(Z) = \{(X, Z, S) \in \text{RK} \mid X = \text{Pa}(Z) \wedge S \in T_s(\sqcup_s(X)) \setminus T_s(\sqcup_s(Z))\}.$$

$K_{\text{RK}}^-(\varphi(X), \varphi(Z))$ ist in $e_{\text{RK}}(Z)$ enthalten, womit

$$\begin{aligned} K_{\text{RK}}^-(\varphi(X), \varphi(Z)) &= K_{\text{RK}}^-((X, X, \sqcup_s(X)), (Z, Z, \sqcup_s(Z))) \\ &= \{(X, Z, S) \mid S \in T_s(\sqcup_s(X)) \setminus T_s(\sqcup_s(Z)) \wedge \sqcup_s(Z) \subset S\} \\ &= \{(X, Z, S) \mid S \in K_s^-(\sqcup_s(X), \sqcup_s(Z)) \vee S = \sqcup_s(X)\} \end{aligned}$$

gilt. Damit ist $|K_{\text{RK}}^-(\varphi(X), \varphi(Z))| = |K_s^-(\sqcup_s(X), \sqcup_s(Z))| + 1 = 1(X, Z)$.

3. Gilt $X \notin \text{DupK}$, so ist

$$e_{\text{RK}}(Z) = \{(X, Z, S) \in \text{RK} \mid X = \text{Pa}(Z) \wedge S \in e_s(Z)\}.$$

$K_{\text{RK}}^-(\varphi(X), \varphi(Z))$ ist in $e_{\text{RK}}(Z)$ enthalten, womit

$$\begin{aligned} K_{\text{RK}}^-(\varphi(X), \varphi(Z)) &= K_{\text{RK}}^-((X, X, \sqcup_s(X)), (Z, Z, \sqcup_s(Z))) \\ &= \{(X, Z, S) \mid S \in e_s(Z) \wedge \sqcup_s(Z) \subset S\} \\ &= \{(X, Z, S) \mid S \in K_s^-(\sqcup_s(X), \sqcup_s(Z))\} \end{aligned}$$

gilt. Damit ist $|K_{\text{RK}}^-(\varphi(X), \varphi(Z))| = |K_s^-(\sqcup_s(X), \sqcup_s(Z))| = 1(X, Z)$.

□

² $X \in \text{DupK}$ folgert aus $\sqcup_s(Z) = \sqcup_s(X)$.

6.1.2.2 Das Maß M_f

Das Maß M_f versucht die Anzahl der realen UGs, welche nicht der Feststellung 1 genügen, abzuschätzen. Alle realen UGs, die nicht der Feststellung 1 genügen, sind Elemente maximaler Teilbäume von UGs in RK , welche wir als *MUTs* (Maximale Unbekannte Teilbäume) bezeichnen. Jedem MUT wird durch das Maß M_f mit der Funktion f ein "Realitätswert" zugeordnet. Dieser Realitätswert schätzt die Anzahl der realen UGs eines MUT ab. Das Maß M_f berechnet sich dann aus der Summe der Realitätswerte aller MUTs.

Beispielsweise kann die Anzahl der realen UGs eines MUTs "grob" durch dessen Blätter abgeschätzt werden. Die Blätter des MUTs beschreiben dabei UGs, welche möglicherweise existieren, aber noch nicht sequenziert worden sind. Wir leiten daraus die heuristische Annahme ab: Je größer die Anzahl der Blätter des MUTs ist, um so unwahrscheinlicher ist es, daß ein Blatt des MUTs noch nicht sequenziert worden ist. Mit f berechnen wir damit für einen MUT eine Anzahl von realen UGs, welche monoton in der Anzahl seiner Blätter fällt.

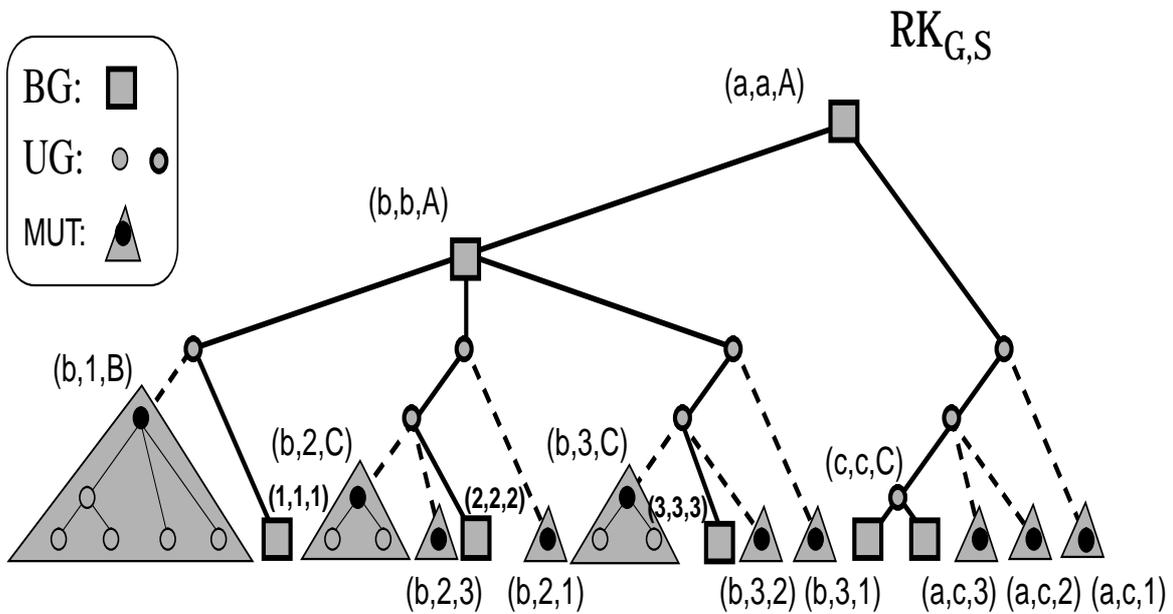


Abbildung 6.2: MUTs in einem konstruktiven Reconciled Tree

In Abbildung 6.2 ist ein MUT in genau einer lokalen Entwicklungsumgebung $e_{RK}(Z)$ mit $Pa(Z) \in \mathcal{G}$ enthalten. Wir identifizieren darin jeden MUT mit seiner Wurzel. Die Wurzeln der MUTs in $e_{RK}(Z)$ "umhüllen" die reduzierte Kette $K_{RK}^-(\varphi(X), \varphi(Z))$. Um diese Wurzeln später auch in S zu beschreiben, definieren wir allgemein die Hülle einer Kette, welche in einem Mengensystem mit Baumordnung enthalten ist.

Definition 6.3 (Hülle einer Kette)

Sei \mathcal{M} ein Mengensystem mit einer Baumordnung und K eine Kette in \mathcal{M} . Dann bezeichnen wir $H_{\mathcal{M}}(K) := \text{Ch}(K) \setminus K$ als die Hülle der Kette K in \mathcal{M} .

Mit dieser Definition ist die Menge aller Wurzeln von MUTs in $e_{\text{RK}}(Z)$ gleich der Menge $H_{\text{RK}}(K_{\text{RK}}^-(\varphi(X), \varphi(Z)))$. In Abbildung 6.2 ergibt sich damit die Menge der Wurzeln aller MUTs durch die Hüllen folgender reduzierter Ketten in RK:

$$\begin{aligned} H_{\text{RK}}(K_{\text{RK}}^-((b, b, A), (1, 1, 1))) &= \{(b, 1, b)\}, \\ H_{\text{RK}}(K_{\text{RK}}^-((b, b, A), (2, 2, 2))) &= \{(b, 2, C), (b, 2, 1)\}, \\ H_{\text{RK}}(K_{\text{RK}}^-((b, b, A), (3, 3, 3))) &= \{(b, 3, C), (b, 3, 2), (b, 3, 1)\}, \\ H_{\text{RK}}(K_{\text{RK}}^-((a, a, A), (c, c, C))) &= \{(a, c, 3), (a, c, 2), (a, c, 1)\} \end{aligned}$$

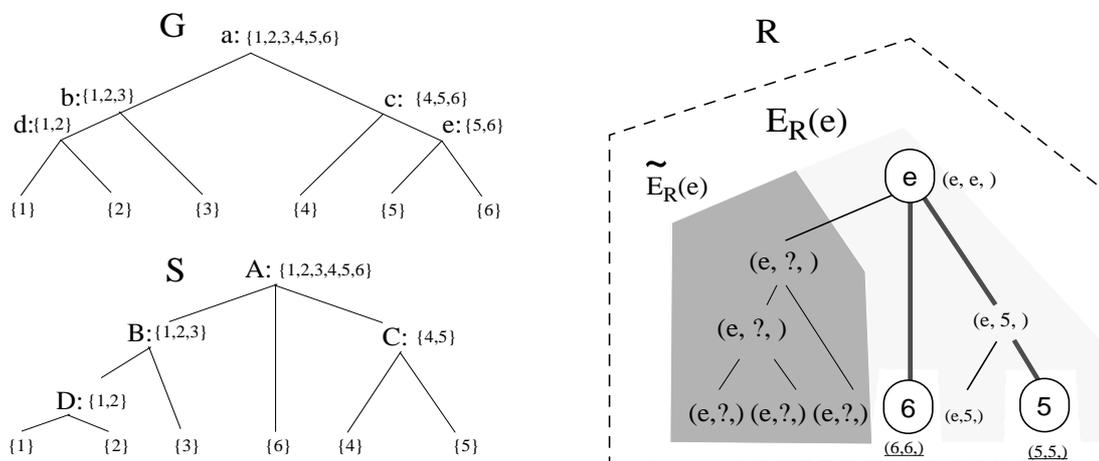


Abbildung 6.3: MUT in der unbestimmten Entwicklungsumgebung $\tilde{E}_{\mathcal{R}}(e)$

Abbildung 6.3 zeigt, daß MUTs nicht nur in lokalen Entwicklungsumgebungen eines Reconciled Tree enthalten sein können, sondern auch in unbestimmten Entwicklungsumgebungen. In Abbildung 6.3 ist die unbestimmte Entwicklungsumgebung $E_{\mathcal{R}}(e)$ gezeigt, die identisch mit dem darin enthaltenen MUT ist. Allgemein kann aber eine unbestimmte Entwicklungsumgebung auch aus mehreren MUTs bestehen.

Wir werden die MUTs einer unbestimmten Entwicklungsumgebung nicht einzeln betrachten, sondern die Gesamtheit aller MUTs direkt durch die unbestimmte Entwicklungsumgebung ausdrücken. Wir gehen davon aus, daß die einzelne Bewertung von MUTs biologisch nur dann einen Sinn macht, wenn wir die beiden bekannten Gene, welche für deren Entwicklung verantwortlich

sind, kennen. Dies ist bei einer lokalen Entwicklungsumgebung der Fall. Beispielsweise ist der MUT in Abbildung 6.2, welcher an $(b, 1, B)$ gewurzelt ist, durch die Entwicklung des bekannten Gens (b, b, A) zu dem bekannten Gen $(1, 1, 1)$ in der lokalen Entwicklungsumgebung $e_{\text{RK}}((1, 1, 1))$ entstanden. In einer unbekanntem Entwicklungsumgebung wissen wir nur, von welchem Gen die Entwicklung des MUTs ausgeht.

Somit bewerten wir für das Maß M_f die MUTs aller lokalen Entwicklungsumgebungen einzeln und die MUTs einer unbestimmten Entwicklungsumgebung zusammen.

Definition 6.4 (M_f)

Sei M eine Menge, $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$ und $f : \text{RK}_{\mathcal{G}, \mathcal{S}} \rightarrow \mathbb{R}$ eine Funktion. Wir definieren das f -Maß wie folgt.

$$M_f(\mathcal{G}, \mathcal{S}) := \sum_{X \in \mathcal{G} \wedge \bar{E}_{\mathcal{R}}(X) \neq \emptyset} f(\varphi(X)) + \sum_{\substack{X \in \mathcal{G}, Z \in \text{Ch}(X), \\ Y'' \in \text{H}(\text{K}_{\text{RK}_{\mathcal{G}, \mathcal{S}}}^-(\varphi(X), \varphi(Z)))}} f(Y'')$$

Ähnlich wie beim Maß M_1 können wir unter einer bestimmten Voraussetzung an f das Maß M_f ohne RK (indirekt) wie folgt berechnen.

Lemma 6.2

Seien $f : \text{RK}_{\mathcal{G}, \mathcal{S}} \rightarrow \mathbb{R}$ und $f' : \mathcal{S} \rightarrow \mathbb{R}$ Funktionen mit $f(Y'') = f'(B(Y''))$ für alle $Y'' \in \text{RK}_{\mathcal{G}, \mathcal{S}}$, dann gilt

$$\begin{aligned} M_f(\mathcal{G}, \mathcal{S}) = & \sum_{X \in \mathcal{G} \wedge \bar{E}_{\mathcal{S}}(X) \neq \emptyset} f'(\bigsqcup_{\mathcal{S}}(X)) + \sum_{\substack{X \in \overline{\text{DupK}}, Z \in \text{Ch}(X), \\ Y' \in \text{H}(\text{K}_{\mathcal{S}}^-(\bigsqcup_{\mathcal{S}}(X), \bigsqcup_{\mathcal{S}}(Z)))}} f'(Y') \\ & + \sum_{\substack{X \in \overline{\text{DupK}}, Z \in \text{Ch}(X), \\ Y' \in \text{H}(\text{K}_{\mathcal{S}}^-(\bigsqcup_{\mathcal{S}}(X), \bigsqcup_{\mathcal{S}}(Z)) \setminus \{\bigsqcup_{\mathcal{S}}(Z)\})}} f'(Y') \end{aligned}$$

Beweis. Seien $X \in \mathcal{G}$ und $Z \in \text{Ch}(\mathcal{G})$. Wir zeigen zunächst

$$1. X \in \overline{\text{DupK}} \Rightarrow B(\text{H}(\text{K}_{\text{RK}}^-(\varphi(X), \varphi(Z)))) = \text{H}(\text{K}_{\mathcal{S}}^-(\bigsqcup_{\mathcal{S}}(X), \bigsqcup_{\mathcal{S}}(Z)))$$

Aus $X \in \overline{\text{DupK}}$ folgt,

$$\text{K}_{\text{RK}}^-(\varphi(X), \varphi(Z)) = \{(X, Z, S) \in \text{RK} \mid S \in \text{K}_{\mathcal{S}}^-(\bigsqcup_{\mathcal{S}}(X), \bigsqcup_{\mathcal{S}}(Z))\}.$$

Damit gilt

$$\begin{aligned}
& \mathbf{B}(\mathbf{H}(\mathbf{K}_{\text{RK}}^-(\varphi(\mathbf{X}), \varphi(\mathbf{Z})))) \\
&= \mathbf{B}(\mathbf{H}(\{(\mathbf{X}, \mathbf{Z}, \mathbf{S}) \in \text{RK} \mid \mathbf{S} \in \mathbf{K}_s^-(\bigsqcup_s(\mathbf{X}), \bigsqcup_s(\mathbf{Z}))\})) \\
&= \mathbf{B}(\{(\mathbf{X}, \mathbf{Z}, \mathbf{S}) \in \text{RK} \mid \mathbf{S} \in \mathbf{H}(\mathbf{K}_s^-(\bigsqcup_s(\mathbf{X}), \bigsqcup_s(\mathbf{Z}))\})) \\
&= \mathbf{H}(\mathbf{K}_s^-(\bigsqcup_s(\mathbf{X}), \bigsqcup_s(\mathbf{Z}))).
\end{aligned}$$

$$2. \mathbf{X} \in \text{DupK} \Rightarrow \mathbf{B}(\mathbf{H}(\mathbf{K}_{\text{RK}}^-(\varphi(\mathbf{X}), \varphi(\mathbf{Z})))) = \mathbf{H}(\mathbf{K}_s(\bigsqcup_s(\mathbf{X}), \bigsqcup_s(\mathbf{Z})) \setminus \{\bigsqcup_s(\mathbf{Z})\})$$

Aus $\mathbf{X} \in \text{DupK}$ folgt

$$\mathbf{K}_{\text{RK}}^-(\varphi(\mathbf{X}), \varphi(\mathbf{Z})) = \{(\mathbf{X}, \mathbf{Z}, \mathbf{S}) \in \text{RK} \mid \mathbf{S} \in \mathbf{K}_s(\bigsqcup_s(\mathbf{X}), \bigsqcup_s(\mathbf{Z})) \setminus \{\varphi(\mathbf{Z})\}\}.$$

Damit gilt

$$\begin{aligned}
& \mathbf{B}(\mathbf{H}(\mathbf{K}_{\text{RK}}^-(\varphi(\mathbf{X}), \varphi(\mathbf{Z})))) \\
&= \mathbf{B}(\mathbf{H}(\{(\mathbf{X}, \mathbf{Z}, \mathbf{S}) \in \text{RK} \mid \mathbf{S} \in \mathbf{K}_s(\bigsqcup_s(\mathbf{X}), \bigsqcup_s(\mathbf{Z})) \setminus \{\varphi(\mathbf{Z})\}\})) \\
&= \mathbf{B}(\{(\mathbf{X}, \mathbf{Z}, \mathbf{S}) \in \text{RK} \mid \mathbf{S} \in \mathbf{H}(\mathbf{K}_s(\bigsqcup_s(\mathbf{X}), \bigsqcup_s(\mathbf{Z})) \setminus \{\varphi(\mathbf{Z})\})\})) \\
&= \mathbf{H}(\mathbf{K}_s(\bigsqcup_s(\mathbf{X}), \bigsqcup_s(\mathbf{Z})) \setminus \{\bigsqcup_s(\mathbf{Z})\}).
\end{aligned}$$

3. Es gilt

$$\sum_{\mathbf{X} \in \mathcal{G} \wedge \tilde{\mathbf{E}}_s(\mathbf{X}) \neq \emptyset} f'(\bigsqcup_s(\mathbf{X})) = \sum_{\mathbf{X} \in \mathcal{G} \wedge \tilde{\mathbf{E}}_{\mathcal{R}}(\mathbf{X}) \neq \emptyset} f(\varphi(\mathbf{X})).$$

Es gilt $\tilde{\mathbf{E}}_s(\mathbf{X}) = \mathbf{B}(\tilde{\mathbf{E}}_{\text{RK}}(\mathbf{X}))$, woraus $\tilde{\mathbf{E}}_s(\mathbf{X}) = \emptyset \Leftrightarrow \tilde{\mathbf{E}}_{\mathcal{R}}(\mathbf{X}) = \emptyset$ folgt.

Da $\mathbf{X} \in \mathcal{G}$ gilt, folgt $\bigsqcup_s(\mathbf{X}) = \mathbf{B}(\varphi(\mathbf{X}))$. Mit $f(\mathbf{Y}'') = f'(\mathbf{B}(\mathbf{Y}''))$ folgt hieraus $f'(\bigsqcup_s(\mathbf{X})) = f'(\mathbf{B}(\varphi(\mathbf{X}))) = f(\varphi(\mathbf{X}))$.

Somit gilt $\tilde{\mathbf{E}}_s(\mathbf{X}) = \emptyset \Leftrightarrow \tilde{\mathbf{E}}_{\mathcal{R}}(\mathbf{X}) = \emptyset$ und $f'(\bigsqcup_s(\mathbf{X})) = f(\varphi(\mathbf{X}))$, woraus die zu zeigende Aussage folgt.

Mit den eben gemachten Aussagen formen wir nun $M_f(\mathcal{G}, \mathcal{S})$ um.

$$\begin{aligned}
M_f(\mathcal{G}, \mathcal{S}) &= \sum_{X \in \mathcal{G} \wedge \tilde{\mathbb{E}}_{\mathcal{R}}(X) \neq \emptyset} f(\varphi(X)) + \sum_{\substack{X \in \mathcal{G}, Z \in \text{Ch}(X), \\ Y'' \in \text{H}(\mathbb{K}_{\mathbb{R}\mathbb{K}}^-(\varphi(X), \varphi(Z)))}} f(Y'') \\
&= \sum_{X \in \mathcal{G} \wedge \tilde{\mathbb{E}}_{\mathcal{S}}(X) \neq \emptyset} f'(\bigsqcup_{\mathcal{S}}(X)) + \sum_{\substack{X \in \mathcal{G}, Z \in \text{Ch}(X), \\ Y'' \in \text{H}(\mathbb{K}_{\mathbb{R}\mathbb{K}}^-(\varphi(X), \varphi(Z)))}} f'(B(Y'')) \\
&= \sum_{X \in \mathcal{G} \wedge \tilde{\mathbb{E}}_{\mathcal{S}}(X) \neq \emptyset} f'(\bigsqcup_{\mathcal{S}}(X)) + \sum_{\substack{X \in \mathcal{G}, Z \in \text{Ch}(X), \\ Y' \in \text{B}(\text{H}(\mathbb{K}_{\mathbb{R}\mathbb{K}}^-(\varphi(X), \varphi(Z))))}} f'(Y') \\
&= \sum_{X \in \mathcal{G} \wedge \tilde{\mathbb{E}}_{\mathcal{S}}(X) \neq \emptyset} f'(\bigsqcup_{\mathcal{S}}(X)) \\
&\quad + \sum_{\substack{X \in \overline{\text{Dup}}\mathbb{K}, Z \in \text{Ch}(X), \\ Y' \in \text{B}(\text{H}(\mathbb{K}_{\mathbb{R}\mathbb{K}}^-(\varphi(X), \varphi(Z))))}} f'(Y') + \sum_{\substack{X \in \text{Dup}\mathbb{K}, Z \in \text{Ch}(X), \\ Y' \in \text{B}(\text{H}(\mathbb{K}_{\mathbb{R}\mathbb{K}}^-(\varphi(X), \varphi(Z))))}} f'(Y') \\
&= \sum_{X \in \mathcal{G} \wedge \tilde{\mathbb{E}}_{\mathcal{S}}(X) \neq \emptyset} f'(\bigsqcup_{\mathcal{S}}(X)) \\
&\quad + \sum_{\substack{X \in \overline{\text{Dup}}\mathbb{K}, Z \in \text{Ch}(X), \\ Y' \in \text{H}(\mathbb{K}_{\mathcal{S}}^-(\bigsqcup_{\mathcal{S}}(X), \bigsqcup_{\mathcal{S}}(Z)))}} f'(Y') + \sum_{\substack{X \in \text{Dup}\mathbb{K}, Z \in \text{Ch}(X), \\ Y' \in \text{H}(\mathbb{K}_{\mathcal{S}}(\bigsqcup_{\mathcal{S}}(X), \bigsqcup_{\mathcal{S}}(Z)) \setminus \{\bigsqcup_{\mathcal{S}}(Z)\})}} f'(Y')
\end{aligned}$$

□

Durch die frei wählbare Funktion f ist das Maß M_f so allgemein, daß sich hierdurch eine Vielzahl von biologischen Erfordernissen formulieren lassen. Unter anderem sind die Maße M_0 und M_1 durch das Maß M_f berechenbar. Für M_1 wählt man dabei f derart, daß für jede Wurzel Y'' eines MUT $f(Y'') = (|\text{Ch}(\text{Pa}(Y''))| - 1)^{-1}$ und sonst $f(Y'') = 0$ ist. Um M_0 zu berechnen, verwendet man die Funktion f für das Maß M_1 . Für jede Wurzel Y'' eines MUT und jede unbestimmte Entwicklungsumgebung addiert man jedoch zusätzlich deren Kardinalität.

Alle bisherigen Maße zum Reconciled Tree lassen sich durch M_f ausdrücken, wie der folgende Abschnitt 6.2 zeigen wird. Jedes in der Allgemeinheit bekannte Maß erfüllt die Bedingung, daß ein f' mit $f(Y'') = f(B(Y''))$ existiert, wodurch dieses nach Lemma 6.2 ohne den Reconciled Tree berechnet werden kann. Das Maß M_f läßt sich auch ohne diese Einschränkung indirekt berechnen, wozu jedoch ein technisch größerer Aufwand erforderlich ist.

6.2 Die Historie der Maße zum Reconciled Tree

In diesem Abschnitt werden wir die chronologische Entwicklung der Maße zum Reconciled Tree $\mathcal{R} \in \text{RC}(\mathcal{G}, \mathcal{S})$ beschreiben. Fast allen Maßen gemeinsam ist, daß sie nur für bifurkierende Phylogenien formuliert und informal beschrieben sind.

Goodman et al. 1979: Das erste Maß wurde von Goodman et al. (1979:138) für bifurkierende Phylogenien formuliert. Als Maß für die Inkonsistenz zwischen \mathcal{G} und \mathcal{R} haben sie die Anzahl von Genduplikationen verwendet, welche sie durch Regeln aus \mathcal{G} und \mathcal{S} bestimmt haben. Diese Regeln entsprechen der Voraussetzung des reduzierten Duplikations-Theorems, Korollar 4.4, für bifurkierende Phylogenien. Damit entspricht das Maß von Goodman et al. der Kardinalität der Menge $\text{DupK}_{\mathcal{G},\mathcal{S}}$ für bifurkierende Phylogenien. Direkt auf dem Reconciled Tree entspricht das Maß von Goodman et al. damit der Kardinalität der Menge Dup_{RK} .

Die Beschreibung der Inkonsistenz zwischen \mathcal{G} und \mathcal{S} durch die Anzahl der Duplikationselemente ist ein sehr "grobes" Maß. Denn die Duplikationen beschreiben jeweils mit dem gleichen Anteil diese Inkonsistenz. Wir haben in Abschnitt 6.1 jedoch gezeigt, daß Duplikationselemente mit einem unterschiedlichen Anteil an UGs die Inkonsistenz zwischen \mathcal{G} und \mathcal{S} beschreiben. Die Anzahl der Duplikationen bildet aber eine untere Schranke für die Anzahl aller möglichen UGs.

Nelson und Platnick 1981: Nelson und Platnick (1981:269) haben als erste UGs verwendet, um ein Maß für die Inkonsistenz zwischen \mathcal{G} und \mathcal{S} zu beschreiben. Dabei haben sie auf bifurkierenden Phylogenien die Knoten gezählt, welche sie zu \mathcal{S} hinzufügen müssen, damit \mathcal{G} einbettbar ist. Page (1990a) hat dieses Maß indirekt durch $|\mathcal{R}| - |\mathcal{G}|$ beschrieben, welches damit der Definition des Maßes M_0 entspricht. Er hat festgestellt, daß damit das Maß M_0 immer gerade Werte liefert (dies gilt offensichtlich nur für bifurkierende Phylogenien). Wichtiger ist jedoch die Feststellung von Page (1994a:65), daß dieses Maß keine direkte biologische Interpretation zuläßt.

Page: R. Page (1988:260) hat die Möglichkeit betrachtet, daß UGs ausgestorben sein können und damit deren Nachfolger nicht existent (irreal) sind. Durch die Untersuchung von MUTs hat er versucht, die realen UGs zu berechnen. Die Idee, als Maß die Anzahl der Blätter aller MUTs zu verwenden, hat Page verworfen (Page,1988), da dies nach seinen Untersuchungen eine Überbewertung

der realen UGs darstellt. Später hat Page (1994a:64) die Anzahl der Wurzeln aller MUTs als Maß verwendet, welches nach seiner Analyse die realen UGs nicht überbewertet und damit sensibel die Inkonsistenz zwischen \mathcal{G} und \mathcal{S} beschreibt. Unter der Voraussetzung, daß nur bifurkierende Phylogenien verwendet werden³ und Feststellung 1 gilt, ist diese Aussage auch nach unserem theoretischen Modell korrekt.

Dazu zeigen wir, daß bei bifurkierenden Phylogenien das Maß von Page dem Maß M_1 entspricht. Gilt Feststellung 1, so ist M_1 eine scharfe untere Schranke für die Anzahl der realen UGs. Damit ist das Maß von Page eine scharfe untere Schranke für die Anzahl der UGs und kann deren Anzahl somit nicht überbewerten. Wir zeigen noch die Äquivalenz des Maßes von Page und dem Maß M_1 für bifurkierende Phylogenien.

Das Maß von Page entspricht dem Maß $M_{f_{\text{Page}}}$, mit $f_{\text{Page}}(X) = 1$, falls X Wurzel eines MUTs ist, und $f_{\text{Page}}(X) = 0$ sonst. Auf bifurkierenden Phylogenien ist dieses Maß gleich dem Maß M_1 , wie folgendes Lemma zeigt.

Lemma 6.3

Sei M eine Menge $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$. Wenn \mathcal{G}, \mathcal{S} bifurkierend sind und $f : \text{RK}_{\mathcal{G}, \mathcal{S}} \rightarrow \{0, 1\}$ eine Funktion mit

$$f(Y'') := \begin{cases} 1 & \exists Z \in \mathcal{G} \setminus \{\text{Ro}(\mathcal{G})\} : Y'' \in H_{\mathcal{R}}(\varphi(\text{Pa}(Z)), \varphi(Z)) \\ 0 & \text{sonst} \end{cases}$$

ist, gilt $M_1 = M_f$.

Beweis. Wenn \mathcal{G} und \mathcal{S} bifurkierend sind, ist auch RK bifurkierend. Also gilt für ein festes aber beliebiges $X \in \mathcal{G}$ und $Z \in \text{Ch}(X)$,

$$|H(K_{\text{RK}}^-(\varphi(X), \varphi(Z)))| = |K_{\text{RK}}^-(\varphi(X), \varphi(Z))|.$$

Also ist

$$\begin{aligned} M_f(\mathcal{G}, \mathcal{S}) &= \sum_{X \in \mathcal{G}} \sum_{Z \in \text{Ch}(X)} \sum_{\substack{Y'' \in \\ H_{\mathcal{R}}(\varphi(X), \varphi(Z))}} f(Y'') \\ &= \sum_{X \in \mathcal{G}} \sum_{Z \in \text{Ch}(X)} |H_{\mathcal{R}}(\varphi(X), \varphi(Z))| \\ &= \sum_{X \in \mathcal{G}} \sum_{Z \in \text{Ch}(X)} |K_{\text{RK}}^-(\varphi(X), \varphi(Z))| \\ &= M_1(\mathcal{G}, \mathcal{S}) \end{aligned}$$

□

³Es ist anzunehmen, daß Page die Aussage auch nur für bifurkierende Phylogenien gemacht hat. Denn er hat in dem Artikel (Page, 1994a) zwar sein Maß für multifurkierende Phylogenien formuliert, jedoch nur bifurkierende Phylogenien analysiert.

Im multifurkierenden Fall ist diese Aussage natürlich nicht korrekt. Auf bifurkierenden Phylogenien gilt offensichtlich $M_1(\mathcal{G}, \mathcal{S}) \leq M_{\text{Page}}(\mathcal{G}, \mathcal{S})$.

B.Mirkin, I.Muchnik, T.Smith 1995: 1995 haben B. Mirkin, I. Muchnik und T. Smith das Problem von inkonsistenten Gen- und Speziesbäumen untersucht. Ohne die Arbeiten von Page und damit den semiformal definierten Reconciled Tree und das Maß M_{Page} von Page zu kennen, haben sie ein Maß M_{Muchnik} für bifurkierende Phylogenien eingeführt (Mirkin et al., 1995). Dieses ist jedoch im wesentlichen zu dem auf bifurkierende Phylogenien beschränkten Page-Maß M_{Page} äquivalent, welches wir zeigen.

Definition 6.5 (Muchnik-Maß)

Sei M eine Menge \mathcal{G}, \mathcal{S} bifurkierende n -Clustersysteme der Menge M , $X \in \mathcal{G}$, $Z \in \text{Ch}(X)$ und

$$l(X, Z) := \begin{cases} |K_S^-(\sqcup_S(X), \sqcup_S(Z))| + 1 & X \in \text{DupK} \wedge \sqcup_S(Z) \neq \sqcup_S(X) \\ |K_S^-(\sqcup_S(X), \sqcup_S(Z))| & X \in \overline{\text{DupK}} \\ 0 & \text{sonst.} \end{cases}$$

Wir bezeichnen $M_{\text{Muchnik}}(\mathcal{G}, \mathcal{S}) := |\text{DupK}| + \sum_{X \in \mathcal{G}, Z \in \text{Ch}(X)} l(X, Z)$ als das Muchnik-Maß für die Inkonsistenz zwischen \mathcal{G} und \mathcal{S} .⁴

Nach Lemma 6.1 ist das Muchnik-Maß ohne die Anzahl der Duplikationselemente $|\text{DupK}|$ die indirekte Definition des direkt definierten Maßes M_1 . Aus Lemma 6.3 folgt, daß für bifurkierende Phylogenien die Maße M_1 und M_{Page} gleich sind. Somit ist für bifurkierende Phylogenien das Muchnik-Maß im wesentlichen die indirekte Definition des Page-Maßes, da $M_{\text{Muchnik}}(\mathcal{G}, \mathcal{S}) - |\text{DupK}| = M_{\text{Page}}$ gilt.

6.3 Die Mirkin-Muchnik-Smith-Vermutung

1995 haben B. Mirkin et al. (1995) das Maß M_{MMS} auf bifurkierenden Phylogenien formuliert, welches auf einer lokalen Betrachtung der Entwicklung aller durch die Duplikationen postulierten Kopien basiert. Obwohl dieses Maß sich von der Konzeption her deutlich von dem Maß M_{Muchnik} unterscheidet, lieferten in der Anwendung die Maße M_{Muchnik} und M_{MMS} immer die gleichen Werte.

⁴Bei einer persönlichen Kommunikation während des *DIMACS-Workshop on Mathematical Hierarchies and Biology 1996* mit B. Mirkin und I. Muchnik stellte sich heraus, daß dieses Maß im wesentlichen von I. Muchnik entwickelt wurde. Daher bezeichnen wir dieses Maß als Muchnik-Maß.

Dies führte zu der Vermutung von Mirkin, Muchnik und Smith, bekannt als *MMS-Vermutung*, welche besagt, daß die Maße M_{Muchnik} und M_{MMS} gleich sind.

Wir werden zunächst in Abschnitt 6.3.1 das Maß M_{MMS} beschreiben und dann in Abschnitt 6.3.2 ein Beispiel für die Äquivalenz der Maße M_{MMS} und M_{Muchnik} zeigen. Abschließend geben wir in Abschnitt 6.3.3 einen kurzen Überblick zu der Struktur des Beweises der MMS-Vermutung in Eulenstein und Vingron (1998).

6.3.1 Das Maß M_{MMS}

Wie das Muchnik-Maß M_{Muchnik} ist das Maß M_{MMS} für bifurkierende Phylogenien definiert und besteht aus zwei additiven Termen. Der erste Term gibt, wie bei dem Maß M_{Muchnik} , die Anzahl der Duplikationselemente $|\text{DupK}|$ an. Der zweite Term beschreibt, vollkommen unterschiedlich zu dem zweiten Term des Maßes M_{Muchnik} , die lokale Entwicklung aller durch Duplikationen postulierten Kopien.

Um genau zu verstehen, welche Elemente der zweite Term berücksichtigt, betrachten wir einen bifurkierenden Genbaum \mathcal{G} , einen bifurkierenden Speziesbaum \mathcal{S} und ein Duplikationselement $X \in \text{DupK}$. Da wir nur bifurkierende Phylogenien betrachten, hat sich das Gen X in die zwei Kopien $X+$ und $X-$ ⁵ dupliziert. Um die Gene, welche sich aus den Kopien $X+$ und $X-$ entwickelt haben zu unterscheiden, bezeichnen wir diese jeweils als $+-$ oder $--$ -Gene.

Das Duplikationselement X hat sich in seiner Host-Spezies $\sqcup_{\mathcal{S}}(X)$ in eine $X+$ und in eine $X-$ Kopie dupliziert. Aus der $+-$ und $--$ -Kopie von X haben sich nun in dem Teilbaum $T_{\mathcal{S}}(\sqcup_{\mathcal{S}}(X))$ die $+-$ und $--$ -Gene entwickelt. Abbildung 6.4 zeigt z.B. die Entwicklung des Duplikationselementes α in dessen Host-Spezies $A (= \sqcup_{\mathcal{S}}(\alpha))$. In jeder Spezies des Teilbaumes $T_{\mathcal{S}}(A)$ hat sich ein $+-$ und $--$ -Gen von α entwickelt. Diese Gene sind jeweils an den Spezies durch zwei Kreise bezeichnet.

Einige der $+-$ und $--$ -Gene in den Blättern des Teilbaumes $T_{\mathcal{S}}(\sqcup_{\mathcal{S}}(X))$ sind existierende Gene, welche durch Blätter von \mathcal{G} dargestellt sind. Da X ein Duplikationselement in \mathcal{G} ist, haben sich die beiden Kinder Z_+ und Z_- von X aus unterschiedlichen Duplikaten entwickelt (dies zeigt ein einfacher Widerspruchsbeweis). Damit sind alle Gene des Teilbaumes $T_{\mathcal{G}}(Z_+)$ $+-$ -Gene und alle

⁵Wir verwenden hier die von Mirkin et al. (1995) verwendete Bezeichnung der beiden Kopien durch $+$ und $-$.

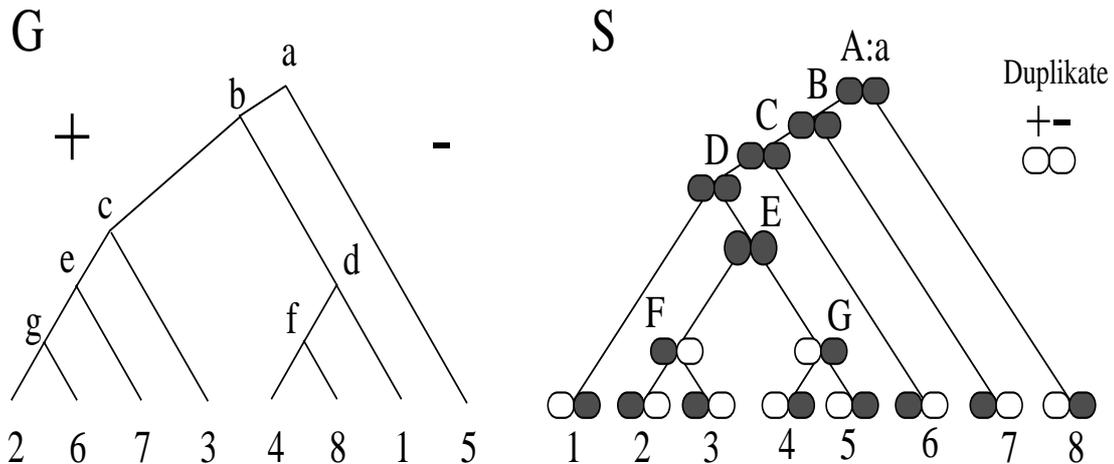


Abbildung 6.4: Die lokale Entwicklung der Duplikate von a

Gene des Teilbaumes $T_g(Z_-)$ --Gene des Gens X. Also bezeichnen die Blätter des Teilbaumes $T_g(Z_+)$ existierende +-Gene und die Blätter des Teilbaumes $T_g(Z_-)$ existierende --Gene von X. Die Host-Spezies der in \mathcal{S} durch Blätter bezeichneten existenten Gene sind die gleichbezeichneten Blätter in \mathcal{S} . Damit ist für jedes Blatt von \mathcal{S} entweder ein existierendes +- oder --Gen von X bekannt. Von einem existenten +-Gen von X mit der Host-Spezies S können wir wiederum auf die Existenz aller +-Gene schließen, deren Host-Spezies auf dem Pfad zwischen der Host-Spezies von X und S liegen. Das gleiche gilt für ein --Gen von X.

Ein Beispiel zeigt Abbildung 6.4. Dort sind die Gene des Teilbaumes $T_g(b)$ die +-Gene und die Gene des Teilbaumes $T_g(5)$ die --Gene von a. Also sind die Blätter 1, 2, 3, 4, 6, 7, 8 +-Gene und das Gen 5 ist ein --Gen. Jedes +- oder --Gen, auf deren Existenz wir dadurch schließen können, daß dieses oder ein Nachfolger von diesem sequenziert worden ist, ist in \mathcal{S} dunkel eingefärbt. Von der Existenz des --Gens 5 von a können wir damit auf die Existenz der --Gene in den Host-Spezies 5, G, E, D, C, B, und A schließen. Entsprechend sind diese dunkel eingefärbt.

Es existieren für eine Spezies vier verschiedene Möglichkeiten, auf die Existenz +- und --Gens von X zu schließen, wonach wir die Menge $T_s(\bigsqcup_{\mathcal{S}}(X))$ wie folgt partitionieren.

1. NC(X): Die Menge der Spezies in $T_s(\bigsqcup_{\mathcal{S}}(X))$, aus der wir jeweils nicht auf die Existenz des +- und --Gens von X schließen können.
2. SC(X): Die Menge der Spezies in $T_s(\bigsqcup_{\mathcal{S}}(X))$, aus der wir jeweils nur auf die Existenz eines +- oder --Gens von X schließen können.

3. $DC(X)$: Die Menge der Spezies in $T_S(\bigsqcup_S(X))$, aus der wir jeweils auf die Existenz des $+$ - und $-$ -Gens von X schließen können.

In Abbildung 6.4 sind die Spezies des Teilbaumes $T_S(\bigsqcup_S(\alpha))$ in die Mengen $NC(X)$, $SC(X)$ und $DC(X)$ durch die Kreise partitioniert. Die Menge $NC(\alpha)$ besteht aus allen Spezies, welche jeweils mit zwei weißen Kreisen markiert sind. Somit gilt $NC(\alpha) = \emptyset$. Die Menge $SC(\alpha)$ besteht aus allen Spezies, welche jeweils mit je einem weißen und einem dunklen Kreis markiert sind. Also gilt $SC(\alpha) = \{1, 2, 3, 4, 5, 6, 7, 8, F\}$. Alle Spezies, die jeweils mit zwei grauen Kreisen markiert sind, bilden die Menge $DC(\alpha) = \{A, B, C, D, E, F\}$.

Formal definieren wir die Menge $DC(X)$ wie folgt:

Definition 6.6

Sei M eine Menge, \mathcal{G}, \mathcal{S} bifurkierende n -Clustersysteme über der Menge M und $X \in \mathcal{G}$, dann ist

$$DC_S(X) := \{Y \in \mathcal{S} \mid Y \subseteq \bigsqcup_S(X) \wedge \forall Z \in \text{Ch}(X) : Z \cap Y \neq \emptyset\}.$$

In Abbildung 6.4 bildet die Menge $DC(\alpha)$ einen zusammenhängenden Teilgraphen und damit einen Baum in $T_S(A)$. Diese Aussage ist offensichtlich auch im allgemeinen Fall gültig. Die Hülle des $DC(\alpha)$ -Baumes gibt alle über eine Kante benachbarten Elemente des $DC(\alpha)$ -Baumes an, welche nicht selbst Element des $DC(\alpha)$ -Baumes sind. Somit besteht die Hülle nur aus Elementen der Menge $NC(\alpha) \cup SC(X)$.

Definition 6.7

Sei M eine Menge, \mathcal{G}, \mathcal{S} bifurkierende n -Clustersysteme über der Menge M und $X \in \mathcal{G}$, dann ist

$$HDC_S(X) := \{Y \in \mathcal{S} \mid \text{Pa}(Y) \in DC(X) \wedge Y \notin DC(X)\}$$

die Hülle der Menge $DC(X)$.

In Abbildung 6.4 besteht die Hülle der Menge $DC(\alpha)$ aus den Elementen 1, F, 4, 5, 6, 7 und 8.

Das Maß M_{MMS} zählt nun in dem zweiten additiven Term für jedes Duplikationselement X in \mathcal{G} die Elemente der Hülle der Menge $HDC(X)$.

Definition 6.8 (Das Maß M_{MMS})

Sei M eine Menge und \mathcal{G}, \mathcal{S} bifurkierende n -Clustersysteme über der Menge M , dann bezeichnen wir

$$M_{MMS}(\mathcal{G}, \mathcal{S}) := |\text{DupK}_{\mathcal{G}, \mathcal{S}}| + \sum_{X \in \text{DupK}_{\mathcal{G}, \mathcal{S}}} |\text{HDC}_{\mathcal{S}}(X)|$$

als das *Mirkin-Muchnik-Smith-Maß* von \mathcal{G} über \mathcal{S} .

6.3.2 Ein Beispiel zu der MMS-Vermutung

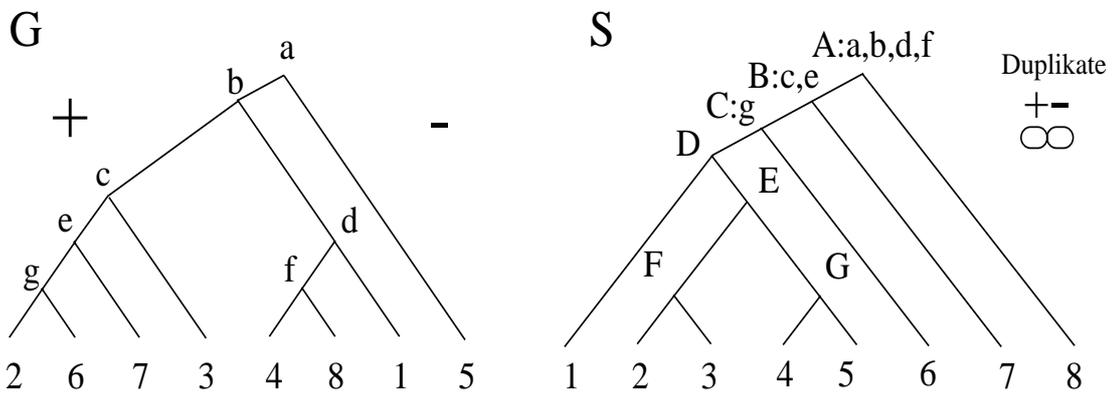


Abbildung 6.5: Genbaum und Speziesbaum

Wir werden für den Genbaum \mathcal{G} und Speziesbaum \mathcal{S} in Abbildung 6.5 zeigen, daß $M_{MMS}(\mathcal{G}, \mathcal{S}) = M_{Muchnik}(\mathcal{G}, \mathcal{S})$ gilt. Um die beiden Maße zu berechnen, bestimmen wir zunächst die Host-Spezies aller Elemente aus \mathcal{G} .

$$\begin{aligned} \sqcup_{\mathcal{S}}(i) &= A, \text{ für jedes } i \in \{a, b, d, f\} & \sqcup_{\mathcal{S}}(j) &= B, \text{ für jedes } j \in \{c, e\} \\ \sqcup_{\mathcal{S}}(i) &= i. \text{ für jedes } i \in \{1, \dots, 8\} \end{aligned}$$

Die Bezeichnungen des Speziesbaumes in Abbildung 6.5 zeigen die Host-Spezies der Elemente aus \mathcal{G} . Damit ist $\text{DupK} = \{a, b, c, d\}$ nach Definition 5.1 (Seite 90). In den beiden folgenden Abschnitten berechnen wir jeweils für das Maß $M_{MMS}(\mathcal{G}, \mathcal{S})$ und $M_{Muchnik}(\mathcal{G}, \mathcal{S})$ den Wert 28, womit $M_{MMS}(\mathcal{G}, \mathcal{S}) = M_{Muchnik}(\mathcal{G}, \mathcal{S})$ gilt.

6.3.2.1 Die Berechnung des Maßes $M_{MMS}(\mathcal{G}, \mathcal{S})$

Um das Maß $M_{MMS}(\mathcal{G}, \mathcal{S})$ zu berechnen, bestimmen wir für jedes Element aus DupK die DC – Mengen.

$$\begin{aligned} DC(a) &= \{A, B, C, D, E, F, G\} & DC(b) &= \{A, B, C, D, E\} \\ DC(c) &= \{B, C, D, E, F\} & DC(d) &= \{A, B, C, D\} \end{aligned}$$

In Abbildung 6.6 sind die Elemente der DC-Mengen durch eine gepunktete Linie gekennzeichnet. Mit der Abbildung 6.6 können wir nun die Elemente der Hüllen abzählen. Damit ergibt sich aus

$$|HDC_S(a)| = 7, |HDC_S(b)| = 6, |HDC_S(c)| = 6, |HDC_S(d)| = 5$$

der Wert $M_{MMS}(\mathcal{G}, \mathcal{S}) = |DupK| + 24 = 28$.

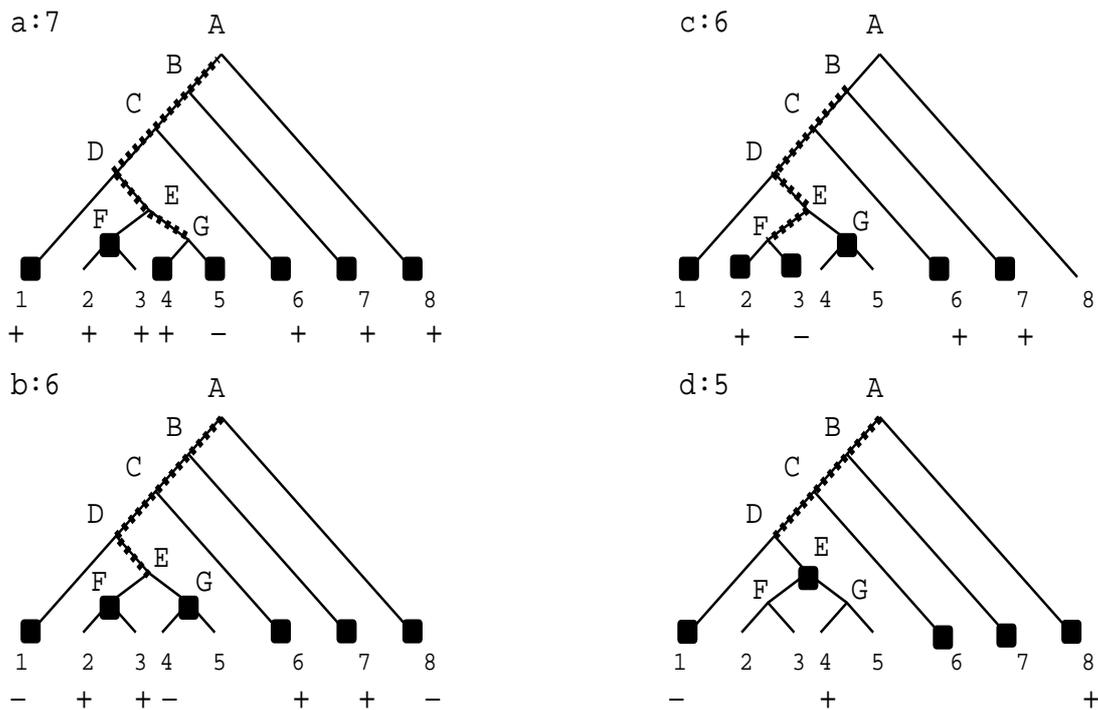


Abbildung 6.6: Die Berechnung des Maßes $M_{MMS}(\mathcal{G}, \mathcal{S})$

6.3.2.2 Die Berechnung des Maßes $M_{\text{Muchnik}}(\mathcal{G}, \mathcal{S})$

Das Muchnik-Maß wird nach Definition 6.5 (Seite 120) wie folgt durch die einzelnen Werte für l berechnet:

$$\begin{array}{ll}
 l(a, b) = 0 & l(a, 5) = |K_{\mathcal{S}}^-(A, 5)| + 1 = 6 \\
 l(b, c) = |K_{\mathcal{S}}^-(A, B)| + 1 = 1 & l(b, d) = 0 \\
 l(c, e) = 0 & l(c, 3) = |K_{\mathcal{S}}^-(B, 3)| + 1 = 5 \\
 l(d, f) = 0 & l(d, 1) = |K_{\mathcal{S}}^-(A, 1)| + 1 = 4 \\
 l(e, g) = |K_{\mathcal{S}}^-(B, C)| = 0 & l(e, 7) = |K_{\mathcal{S}}^-(B, 7)| = 0 \\
 l(f, 4) = |K_{\mathcal{S}}^-(A, 4)| = 5 & l(f, 8) = |K_{\mathcal{S}}^-(A, 8)| = 0 \\
 l(g, 2) = |K_{\mathcal{S}}^-(C, 2)| = 3 & l(g, 6) = |K_{\mathcal{S}}^-(C, 6)| = 0
 \end{array}$$

Alle diese Werte summieren sich zu 24 auf. Für das Muchnik-Maß zu \mathcal{G} über \mathcal{S} ergibt sich somit $M_{\text{Muchnik}}(\mathcal{G}, \mathcal{S}) = |\text{DupK}| + 24 = 28$.

6.3.3 Eine Idee zum Beweis der MMS-Vermutung

Der formale Beweis ist sehr komplex, daher zeigen wir die wesentliche Idee des Beweises an dem Beispiel für den Genbaum \mathcal{G} und den Speziesbaum \mathcal{S} aus Abbildung 6.5. Um die von beiden Maßen gezählten Knoten vergleichen zu können, werden wir Kanten anstatt Knoten zählen. Dazu geben wir jeweils eine äquivalente Umformulierung der Maße an.

Umformulierung von M_{MMS} : Für jedes Element Y , welches von dem Maß M_{MMS} gezählt wird, zählen wir statt dessen die zugehörige Kante $(\text{Pa}(Y), Y)$. Zum Beispiel wird in Abbildung 6.6 das Element 1 in der Hülle $\text{HDC}(a)$ durch das Maß M_{MMS} gezählt. Statt dessen zählen wir die Kante $(D, 1)$.

Umformulierung von M_{Muchnik} : Sei X ein Element, welches durch das Maß M_{Muchnik} gezählt wird. Damit wird X für eine Kante (a, b) in \mathcal{G} durch $l(a, b)$ gezählt. Also ist X Element der reduzierten Kette $K_{\mathcal{S}}^-(\sqcup_{\mathcal{S}}(a), \sqcup_{\mathcal{S}}(b))$ und damit Element der Kette $K_{\mathcal{S}}(\sqcup_{\mathcal{S}}(a), \sqcup_{\mathcal{S}}(b))$. Das Element X besitzt zwei Kinder. Ein Kind ist Element der Kette $K_{\mathcal{S}}(\sqcup_{\mathcal{S}}(a), \sqcup_{\mathcal{S}}(b))$, das andere Kind Y ist kein Element dieser Kette. Anstatt das Element X für das Maß M_{MMS} zu zählen, werden wir nun die Kante (X, Y) zählen. Zum Beispiel zählt das Maß M_{Muchnik} das Element $B \in \mathcal{S}$ durch $l(a, 5)$ für die Kante $(a, 5)$ in \mathcal{G} . Das Element B und dessen Kind C sind Elemente der Kette $K_{\mathcal{S}}^-(\sqcup_{\mathcal{S}}(a), \sqcup_{\mathcal{S}}(5))$. Das Kind 7 ist damit kein Element

dieser Kette. Somit zählen wir die Kante $(B, 7)$ anstatt des Elementes B . Anstatt bei Berechnung von $l(X, Z)$ im Fall $X \in \text{DupK} \wedge \sqcup_S(Z) \neq \sqcup_S(X)$ den Wert 1 zu addieren, zählen wir die Kante $(\sqcup_S(X), Y)$ mit $Y \notin K_S^-(\sqcup_S(X), \sqcup_S(Z))$. Zum Beispiel wird für $l(a, 5)$ der Wert $|K_S^-(\sqcup_S(a), \sqcup_S(5))| + 1$ berechnet. Für den Wert 1 zählen wir die Kante $(\sqcup_S(a), 8)$, da 8 kein Element der reduzierten Kette $K_S^-(\sqcup_S(a), \sqcup_S(5))$ ist.

	a	b	C	d	e	f	g	$\sum_X 1$	$\sum_O 1$
(D,E)				X O				1	1
(E,F)	X O	X				O		2	2
(E,G)		X	X O				O	2	2
(D,1)	X O	X	X O	X		O	O	4	4
(F,2)			X O					1	1
(F,3)			X				O	1	1
(G,4)	X O							1	1
(G,5)	X					O		1	1
(C,6)	X O	X	X O	X O		O		4	4
(B,7)	X O	X	X O	X O		O		4	4
(A,8)	X O	X O		X O				3	3

Tabelle 6.1: Kanten in \mathcal{S} , welche die Maße $M_{\text{MMS}}(X)$ und $M_{\text{Muchnik}}(O)$ zählen

Die obige Tabelle 6.1 zeigt die durch das Maß M_{MMS} und das Maß M_{Muchnik} gezählten Kanten in \mathcal{S} . Dabei bedeutet ein Eintrag X oder O in der Spalte I und Zeile J folgendes:

1. Der Eintrag X bedeutet, daß die Kante J in der Hülle von $\text{HDC}(I)$ durch das Maß $M_{\text{MMS}}(\mathcal{G}, \mathcal{S})$ gezählt wird.
2. Der Eintrag O bedeutet, daß die Kante J durch das Maß M_{Muchnik} gezählt wird. Das Maß M_{Muchnik} zählt dabei J durch $l(I, Z)$ für ein $Z \in \text{Ch}(I)$.

Tabelle 6.1 zeigt an dem Beispiel aus Abbildung 6.5 die Beweisidee. Von jedem der beiden Maße wird eine beliebige Kante in \mathcal{S} mit dem gleichen Wert gezählt. Was hier an einem Beispiel gezeigt ist, haben wir für den allgemeinen Fall gezeigt und damit die MMS-Vermutung bewiesen.

6.4 Ausblick

Wir haben die bisherigen Maße zum Reconciled Tree dargestellt, deren biologische Aussagekraft und Äquivalenzen aufgezeigt. Dabei haben wir festgestellt, daß die sensibleren biologischen Maße, wie M_{Muehnik} , M_{MMS} und $M_{\text{f}_{\text{page}}}$ im wesentlichen mit dem Maß M_1 auf bifurkierenden Phylogenien identisch sind. Aufgrund der unabhängigen Entwicklung der Maße M_{Muehnik} , M_{MMS} und $M_{\text{f}_{\text{page}}}$ läßt deren Äquivalenz eine biologische Signifikanz dieser Maße vermuten.

Im wesentlichen sind alle historischen Maße zum Reconciled Tree für bifurkierenden Phylogenien definiert worden. Multifurkierende Phylogenien besitzen jedoch ein wesentlich größeres Aussagespektrum⁶, da durch diese mögliche Unsicherheiten in dem Wissen über deren Struktur dargestellt werden können. Ist beispielsweise über die Entwicklung einer Menge von existierenden Genen nichts bekannt, so können diese als Kinder einer Wurzel dargestellt werden. Bei einer bifurkierenden Phylogenie wäre man gezwungen, die exakte Phylogenie dieser Gene zu erraten. Die in Abschnitt 6 gezeigten Maße sind alle für multifurkierende Phylogenien konzipiert, jedoch noch nicht angewandt worden. Diese Maße, im besonderen das Maß M_f , sollen als theoretische Grundlage für die Definition weiterer, aus der konkreten Anwendung abgeleiteter, Maße des Reconciled Tree dienen.

Robustifizierung

Zu der theoretischen Betrachtung der Inkonsistenz von Gen- und Speziesbaum haben wir den zu analysierenden Gen- und Speziesbaum als korrekt vorausgesetzt. In der Praxis müssen wir aber davon ausgehen, daß diese Bäume fehlerhaft sind. Ein kleiner Fehler in Gen- oder Speziesbaum kann zu einer maximalen Inkonsistenz zwischen diesen Phylogenien führen. Um nicht durch kleine Ungenauigkeiten in Gen- oder Speziesbaum sinnlose Schlußfolgerungen zu erhalten, ist von Vingron (persönliche Kommunikation) die *Robustifizierung* des Reconciled Tree vorgeschlagen worden. Unter der Robustifizierung des Reconciled Tree sind Analysemethoden zu verstehen, welche den Reconciled Tree gegen unsinnige Schlußfolgerungen "robust" machen. Solche Analysemethoden existieren bisher nur im Ansatz. Vorstellbar ist, hier die "Qualität" der Duplikationen in die Kriterien der Rekonstruktion des Gen- oder Speziesbaumes direkt mit einfließen zu lassen. Dies kann z.B. durch die Kombination von phylogenetischen Maßen mit einem Maß zum Reconciled Tree realisiert werden.

⁶Bifurkierende Phylogenien sind multifurkierende Phylogenien.

Das phylogenetische Maß beschreibt, wie “teuer” eine lokale Änderung, z.B. in dem Genbaum, ist. Das Maß zum Reconciled Tree gibt an, wie sich diese lokale Änderung auf die Inkonsistenz zwischen Gen- und Speziesbaum auswirkt. Hierbei ist dann ein Optimum zu finden.

Verteilung der Maße

Um die Maße zum Reconciled Tree besser interpretieren zu können, ist es notwendig, deren Verteilung auf geordneten Paaren aus Gen- und Speziesbaum zu kennen, welche durch einen Evolutions-Prozeß entstanden sind. Möglicherweise existieren Eigenschaften der Maße, welche sich invariant gegenüber der Verteilung der Paare aus Gen- und Speziesbaum verhalten. Hierbei erschweren insbesondere die meist fehlenden Struktureigenschaften, wie z.B. die Symmetrie von den gezeigten sensiblen Maßen zum Reconciled Tree, die Analyse. Beispielsweise haben wir für den Erwartungswert von M_1 , bei einer Gleichverteilung der Phylogenien, bisher nur eine komplexe rekursive Formel gefunden. Dabei ist anzunehmen, daß die Aussagekraft dieses Erwartungswertes nur einen Anhaltspunkt für den wahren Erwartungswert liefert, da es unwahrscheinlich ist, daß die Evolution die Phylogenien gleich verteilt hat.

Normierung

Wir wollen mit einem Maß M zum Reconciled Tree nicht nur die Inkonsistenz von Gen- und Speziesbaum $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$ für eine feste Menge M vergleichen. Um Gen- und Speziesbaum aus Mengen $\mathfrak{N}(M_i)$ und $\mathfrak{N}(M_j)$ mit $M_i \neq M_j$ zu vergleichen, benötigen wir geeignete Skalierungsfaktoren zum normalisieren des Maßes M .

Lokale und globale Maße

Die Entwicklung der Maße zum Reconciled Tree ist auch in einem lokaleren oder globaleren Rahmen zu sehen.

Lokal ist hierbei die exakte Analyse der Inkonsistenz eines bestimmten Bereiches des Genbaumes zu einem bestimmten Bereich des Speziesbaumes zu verstehen. Hierunter ist z.B. die Entwicklung von Maßen für die in Abschnitt 5.3 erwähnten Mixed Trees zu verstehen. Der Teil eines Mixed Tree, welcher aufgelöste Kanten des Genbaumes darstellt, ist konsistent zu dem Speziesbaum. Die “Kosten” der Aufhebung dieser lokalen Inkonsistenzen sind dabei durch ein geeignetes Maß zu beschreiben. Insbesondere sollten diese Maße unter dem Gesichtspunkt der Robustifizierung konstruiert werden, da lokale Veränderungen des Genbaumes durch den Mixed Tree gut analysierbar sind.

Globale Maße beschreiben die Inkonsistenz einer Menge von Genbäumen zu ei-

nem Speziesbaum. Beispielsweise kann man mit einem solchen Maß die Menge der Speziesbäume suchen, welche unter diesem Maß eine minimale Inkonsistenz besitzen. Damit ließe sich dann aus einer Menge von Genbäumen ein Speziesbaum unter dem Gesichtspunkt der Genduplikation konstruieren. Page et al. (1997) haben auf diese Weise Speziesbäume konstruiert. Li et al. (1998) haben für ein globales Maß gezeigt, daß dessen Berechnung NP-vollständig ist.

Das theoretische Fundament für die Maße zum Reconciled Tree ist vorhanden. Diese Maße müssen jetzt an die Praxis adaptiert werden, um für den Biologen eine effektive Methode zu der phylogenetischen Analyse von Genduplikationen zu implementieren.

Kapitel 7

Algorithmen zur Vorhersage von Genduplikationen

Sei M eine Menge, $n := |M|$ und $\mathcal{G}, \mathcal{S} \in \mathfrak{N}(M)$, wobei \mathcal{G} ein Genbaum und \mathcal{S} ein Speziesbaum ist. In diesem Kapitel beschreiben wir Algorithmen, mit denen sich alle Genduplikationen in \mathcal{G} bezüglich \mathcal{S} und das Maß $M_1(\mathcal{G}, \mathcal{S})$ berechnen lassen. Das Maß M_1 ist, wie in Abschnitt 6.2 gezeigt, zu den aus der Literatur bekannten Maßen von Page (1994a), Mirkin et. al. (1996) sowie Mirkin et. al. (1995)¹ äquivalent und daher von besonderem Interesse.

Bisher gab es keine Algorithmen zu der Berechnung von Genduplikationen und Maßen auf multifurkierenden Phylogenien, da die hierzu notwendigen strukturellen Kenntnisse fehlten. Der Algorithmus von Page berechnet Genduplikationen auf bifurkierenden Phylogenien (Page, 1994). Für diesen Algorithmus ist von Page weder dessen Korrektheit noch seine Laufzeit gezeigt worden. Eine einfache Analyse ergibt eine Laufzeit von $O(n^2)$, und dessen Korrektheit wird erst mit Korollar 4.4 offensichtlich. Zhang (1997) zeigt durch eine Reduktion auf die Berechnung von kleinsten gemeinsamen Vorfahren in \mathcal{S} , daß alle Genduplikationen in \mathcal{G} unter gewissen technischen Voraussetzungen in der Zeit $O(n)$ zu berechnen sind.

Um Genduplikationen auf multifurkierenden Phylogenien vorherzusagen, berechnen wir die Duplikationselemente in \mathcal{G} . Ein Element $X \in \mathcal{G}$ ist nach Definition 5.1 genau dann ein Duplikationselement, wenn die folgende Aussage nicht für alle $Z_i, Z_j \in \text{Ch}(X)$ mit $i \neq j$ gilt.

$$\bigsqcup_{\mathcal{S}}(Z_i) \subset \bigsqcup_{\mathcal{S}}(X) \text{ und } \bigsqcup_{\mathcal{S}}(Z_i) \sqcup \bigsqcup_{\mathcal{S}}(Z_j) = \bigsqcup_{\mathcal{S}}(X)$$

¹Die Ordnung der Literaturangaben entspricht der historischen Entstehung der Maße und nicht deren erstmaliger Veröffentlichung.

Die Bestimmung der Duplikationselemente unterteilt sich demnach in die beiden folgenden Probleme.

1. *Dup-Problem*: Bestimme bei gelöstem LCA-Problem für \mathcal{G} und \mathcal{S} alle Duplikationselemente in \mathcal{G} .
2. *LCA-Problem*: Bestimme für den fest vorgegebenen Genbaum \mathcal{G} das Element $\sqcup_{\mathcal{S}}(X)$ für jedes $X \in \mathcal{G}$.

Zuerst zeigen wir in Abschnitt 7.1, daß das Dup-Problem in der Zeit $O(n)$ lösbar ist. In Abschnitt 7.2 geben wir dann einen direkt auf das LCA-Problem zugeschnittenen Algorithmus, den LCA-Algorithmus, an. Mit diesem ist das LCA-Problem in der Zeit $O(n \alpha(n, n))$ lösbar. Dabei ist $\alpha := \min\{i \geq 1 \mid A(i, \lfloor n/i \rfloor) > \lg(n)\}$ die *inverse Ackermann-Funktion*, und A bezeichnet die Ackermann-Funktion (Cormen et al, 1996). Somit sind alle Duplikationselemente von \mathcal{G} in der Zeit $O(n \alpha(n, n))$ bestimmbar. In jeder vorstellbaren Anwendung ist $\alpha(n, n) \leq 4$, womit wir die Laufzeit des Algorithmus in der Praxis als linear ansehen können.

Unter gewissen technischen Randbedingungen ist der LCA-Algorithmus einer der asymptotisch schnellsten Algorithmen, um das LCA-Problem zu lösen. Das LCA-Problem ist aber auch in linearer Zeit auf das Problem der Berechnung von kleinsten oberen Schranken in \mathcal{S} reduzierbar. Für diese Berechnung existieren bereits schnelle Algorithmen. Asymptotisch ergibt sich hierdurch aber bisher bestenfalls eine zu dem LCA-Algorithmus gleiche Laufzeit. Für die Anwendung sind die bei der asymptotischen Laufzeit vernachlässigten konstanten Faktoren von Bedeutung. Da der LCA-Algorithmus direkt auf das Anwendungsproblem zugeschnitten ist, besitzt dessen Laufzeit einen kleineren konstanten Faktor als die Laufzeit der reduzierten Variante. Die Reduktion sowie Algorithmen zur Lösung des reduzierten Problem es zeigen wir in Abschnitt 7.3.

In Abschnitt 7.4 zeigen wir die einfache Berechnung des Maßes M_1 in der Zeit $O(n)$, wenn das Dup-Problem gelöst wurde. Somit ist auch das Maß M_1 in der Zeit $O(n \alpha(n, n))$ berechenbar.

7.1 Eine Lösung des Dup-Problemes

Sei $X \in \mathcal{G}$. Die Prozedur `DUPLIKATIONSVORHERSAGE` entscheidet für das Element X , ob es ein Duplikationselement in \mathcal{G} bezüglich \mathcal{S} ist. Dabei setzen wir voraus, daß für jedes $Y \in \mathcal{G}$ das Element $\sqcup_{\mathcal{S}}(Y)$ bereits bekannt ist. Als Ausgabe

wird die Variable *Duplikationselement*(X) verwendet. Gilt *Duplikationselement*(X) = TRUE, so ist X ein Duplikationselement. Sonst gilt *Duplikationselement*(X) = FALSE und X ist kein Duplikationselement.

Aus Definition 5.1 folgt, daß X genau dann ein Duplikationselement ist, wenn entweder die Aussage $\sqcup_s(Z) = \sqcup_s(X)$ für ein $Z \in \text{Ch}(X)$ oder die Aussage $\sqcup_s(Z) \sqcup \sqcup_s(Z') \subset \sqcup_s(X)$ für ungleiche Kinder $Z, Z' \in \text{Ch}(X)$ gilt. Die Prozedur DUPLIKATIONSVORHERSAGE untersucht diese beiden Aussagen für jedes Kind $Z \in \text{Ch}(X)$ und entscheidet dadurch, ob X ein Duplikationselement ist.

Die Aussage $\sqcup_s(Z) = \sqcup_s(X)$ überprüft die Prozedur DUPLIKATIONSVORHERSAGE in Zeile 4. Gilt die Aussage, so ist X ein Duplikationselement und der Variablen *Duplikationselement*(X) wird in Zeile 5 der Wert TRUE zugeordnet. Sonst wird die initiale Zuordnung *Duplikationselement*(X) = FALSE beibehalten.

Die Aussage $\sqcup_s(Z) \sqcup \sqcup_s(Z') \subset \sqcup_s(X)$ überprüft die Prozedur *Duplikationselement* in den Zeilen 6 bis 11. Dazu besucht diese für das Kind Z die Elemente des Pfades von $\sqcup_s(Z)$ nach $\sqcup_s(X)$, solange diese nicht schon vorher von einem anderen Kind $Z' \in \text{Ch}(X)$ aus besucht worden sind.

Ob ein Element K des Pfades schon besucht wurde, wird durch dessen Variable *visited*(K) festgestellt. Gilt *visited*(K) = TRUE, so ist K vorher besucht worden. Sonst gilt die initiale Zuordnung *visited*(K) = FALSE, und K ist vorher noch nicht besucht worden.

Besucht die Prozedur ein Element K' ein zweites Mal, so ist K' bereits von einem anderen Kind $Z' \in \text{Ch}(X)$ besucht worden. Also existiert ein zu Z ungleiches Kind $Z' \in \text{Ch}(X)$, für welches $\sqcup_s(Z) \sqcup \sqcup_s(Z') = K'$ gilt. Gilt zusätzlich $K' \subset \sqcup_s(X)$, so ist X ein Duplikationselement. Die Prozedur überprüft die Aussage $K' \subset \sqcup_s(X)$ in Zeile 10. Gilt die Aussage, so wird der Variablen *Duplikationselement*(X) in Zeile 11 der Wert TRUE zugeordnet.

INIT(G)

01 Initialisiere

02 *visit*(K) \leftarrow FALSE für jedes $K \in E_s(G)$ und

03 *Duplikationselement*(G) \leftarrow FALSE

```

DUPLIKATIONSVORHERSAGE(G)
00 INIT(G)
01 visited( $\sqcup_S(G)$ )  $\leftarrow$  TRUE
02 while NEXT-CHILD( $\sqcup_S(G)$ )  $\neq$  NIL and Duplikationselement(G) =FALSE
03     do child  $\leftarrow$  NEXT-CHILD( $\sqcup_S(G)$ )
04         if  $\sqcup_S(\textit{child}) = \sqcup_S(G)$ 
05             then Duplikationselement(G)  $\leftarrow$  TRUE
06             else while visited(Pa(child))  $\neq$  FALSE
07                 do visited( $\sqcup_S(G)$ )  $\leftarrow$  TRUE
08                     child  $\leftarrow$  Pa(child)
09                     visited(child)  $\leftarrow$  TRUE
10             if child  $\neq \sqcup_S(G)$ 
11                 then Duplikationselement(G)  $\leftarrow$  TRUE

```

Der Zeitaufwand für diesen Algorithmus bei Eingabe X beträgt $O(E_S(X))$, da nur Elemente der Entwicklungsumgebung $E_S(X)$ höchstens einmal besucht werden. Die Entwicklungsumgebungen in S für alle Elemente aus \mathcal{G} partitionieren S . Damit beträgt die Laufzeit $O(|S|)$ für die Berechnung aller Duplikationselemente in \mathcal{G} . Da $S \in \mathfrak{N}(M)$ gilt, ist n die Anzahl der Blätter von S . Aus $|S| \leq 2n - 1$ und $O(|S|)$ erhalten wir die Laufzeit $O(n)$ für die Bestimmung aller Duplikationselemente.

7.2 Eine direkte Lösung des LCA-Problemes

In diesem Abschnitt beschreiben wir den LCA-Algorithmus, welcher direkt auf das LCA-Problem zugeschnitten ist. Eingabe des Algorithmus sind der Genbaum \mathcal{G} und Speziesbaum S . Ausgegeben wird das Element $\sqcup_S(X)$ in der Variablen $\coprod_S(X)$ für jedes $X \in \mathcal{G}$.

In dem folgenden Abschnitt 7.2.1 beschreiben wir den LCA-Algorithmus und verifizieren diesen in Abschnitt 7.2.2. Dessen Laufzeit zeigen wir dann in Abschnitt 7.2.3.

7.2.1 Der LCA-Algorithmus

Der LCA-Algorithmus besteht aus zwei Prozeduren, der rekursiven Prozedur SPECIES TREE-DFS und der iterativen Prozedur GENETREE-WALK. Die Prozedur SPECIES TREE-DFS besucht nur Elemente des Speziesbaumes S , die Prozedur GENETREE-WALK nur Elemente des Genbaumes \mathcal{G} .

7.2.1.1 Die Prozedur SPECIES TREE-DFS

Die Berechnung beginnt mit dem Aufruf der Prozedur SPECIES TREE-DFS($\text{Ro}(S)$), welche den Spezies-Baum S postorder durchläuft. Während des DFS-Durchlaufes werden Körbe, sogenannte Baskets, durch die Operation CREATE-BASKET in Zeile 10 für jedes Blatt erzeugt, wenn dieses besucht wird.

Baskets beinhalten nur Elemente aus \mathcal{G} und sind nur Elementen aus S zugeordnet. Einem Element aus S ist höchstens ein Basket zugeordnet, und jedes Element aus \mathcal{G} ist in höchstens einem Basket beinhaltet. Den Basket, welcher einem Element $S \in S$ zugeordnet ist, bezeichnen wir als $\text{basket}(S)$. Wird eine Kante während des DFS-Durchlaufes von S in Richtung der Wurzel $\text{Ro}(S)$ von Knoten Y zu Knoten X durchlaufen, so wird $\text{basket}(Y)$ dem Knoten X "zugeewiesen". Wenn X noch keinen Basket besitzt, erfolgt dies durch die Operation MOVE-BASKET in Zeile 4, welche $\text{basket}(Y)$ dem Element Y zuordnet. Sonst wird durch die Operation MERGE-BASKET der Inhalt von $\text{basket}(Y)$ dem Inhalt von $\text{basket}(X)$ zugefügt.

Erreicht der Durchlauf ein Blatt $L \in S$ (Zeile 11), so wird ein Basket erzeugt, welcher L zugeordnet ist. Für das Blatt L in S ist das Blatt $\coprod_S^{-1}(L)$ in \mathcal{G} durch die Initialisierung bekannt. Danach sind die Blätter von \mathcal{G} und S identisch, womit das Blatt $\coprod_S^{-1}(L)$ in \mathcal{G} dem Blatt L in S entspricht. Mit dem Element $\coprod_S^{-1}(L)$ wird nun der für L erzeugte Basket gefüllt. Die Erzeugung des Baskets und dessen initiale Füllung erfolgt durch die Operation CREATE-BASKET.

Im folgenden ist die Prozedur SPECIES TREE-DFS und die von dieser Prozedur verwendeten Operationen für Baskets angegeben. Seien $X, Y \in S$ und $G \in \mathcal{G}$.

```

SPECIES TREE-DFS( $S$ )
01 if IS-NOT-LEAF( $S$ )
02   then  $child \leftarrow \text{NEXT-CHILD}(S)$ 
03       SPECIES TREE-DFS( $child$ )
04       MOVE-BASKET( $child, S$ )
05        $child \leftarrow \text{NEXT-CHILD}(S)$ 
06       while  $child \neq \text{NIL}$ 
07           do SPECIES TREE-DFS( $child$ )
08               MERGE-BASKET( $child, S$ )
09                $child \leftarrow \text{NEXT-CHILD}(S)$ 
10   else CREATE-BASKET( $\coprod_S^{-1}(S), S$ )
11       GENETREE-WALK( $S$ )

```

1. CREATE-BASKET(G, X) erzeugt einen Basket, welcher dem Element X zugeordnet ist und initial das Element G beinhaltet.
2. MOVE-BASKET(Y, X) ordnet basket(Y) dem Element X zu.
3. MERGE-BASKET(Y, X) fügt die Elemente von basket(Y) den Elementen von basket(X) hinzu und löscht basket(Y).

7.2.1.2 Die Prozedur GENETREE-WALK

Seien $\{l_0\}, \dots, \{l_n\}$ (es gilt $n = |M|$) die Blätter des Speziesbaumes \mathcal{S} . Besucht ein DFS-Durchlauf gerade das Blatt l_i und hat zuvor die Blätter $\{l_0\}, \dots, \{l_{i-1}\}$ besucht, berechnet die Prozedur GENETREE-WALK($\coprod_{\mathcal{S}}^{-1}(\{l_i\})$) das Element $\coprod_{\mathcal{S}}(X)$ für jedes $X \in \mathcal{G}$ mit $X \subseteq \{l_0, \dots, l_{i-1}, l_i\}$ und $l_i \in X$.

Dazu durchläuft die Prozedur GENETREE-WALK den Genbaum \mathcal{G} von dessen Blatt $\{l_i\}$ in Richtung $\text{Ro}(\mathcal{G})$. Der Durchlauf terminiert genau dann an einem Knoten $Y \in \mathcal{G}$, wenn dessen Elternknoten $\text{Pa}(Y)$ noch nicht über alle seine Kinder durch Aufrufe von GENETREE-WALK besucht worden ist. Dies wird in Zeile 1 durch die Variable *visited-childrens*($\text{Pa}(Y)$) überprüft, welche die Anzahl seiner bisher besuchten Kinder angibt.

Terminiert der Durchlauf an dem Knoten Y , so ist das Kind Y von $\text{Pa}(Y)$ besucht worden, womit der Wert *visited-childrens*($\text{Pa}(Y)$) in Zeile 8 um eins erhöht wird.

Terminiert der Durchlauf nicht an einem besuchten Knoten $\text{Pa}(Y)$, so sind alle Kinder von $\text{Pa}(Y)$ besucht worden. Damit ist $\text{Pa}(Y) \subseteq \{l_0, \dots, l_{i-1}, l_i\}$ und $l_i \in \text{Pa}(Y)$. Außerdem sind nun alle Kinder von $\text{Pa}(Y)$ in Baskets enthalten, womit das Element $\coprod_{\mathcal{S}}(\text{Pa}(Y))$ in den Zeilen 3 bis 5 berechnet werden kann. Mit der Operation FIND-BASKET werden dazu jeweils die Baskets gesucht, in welchen die Kinder von $\text{Pa}(Y)$ enthalten sind. Unter den Knoten in \mathcal{S} , welche diesen Baskets zugeordnet sind, wird ein Knoten S mit geringster Tiefe ausgewählt und der Variablen $\coprod_{\mathcal{S}}(\text{Pa}(Y))$ zugewiesen. Wir werden in Abschnitt 7.2.2 noch zeigen, daß $\coprod_{\mathcal{S}}(\text{Pa}(Y)) = \sqcup_{\mathcal{S}}(\text{Pa}(Y))$ gilt, womit auch S eindeutig bestimmt ist.

Um die Elemente $\sqcup_{\mathcal{S}}(K)$ für jedes Element K auf dem Weg von $\text{Pa}(Y)$ zur Wurzel $\text{Ro}(\text{Pa}(G))$ bestimmen zu können, wird mit der Operation INSERT-INTO-BASKET der Knoten $\text{Pa}(Y)$ in basket(S) eingefügt.

Im folgenden ist die Prozedur `GENETREE-WALK` und die von dieser Prozedur verwendeten Operationen für Baskets angegeben. Seien $G \in \mathcal{G}$ und $S \in \mathcal{S}$.

`GENETREE-WALK(G)`

```

01 while  $G \neq \text{Ro}(\mathcal{G})$  and  $|\text{Pa}(G)| = \text{visited-children}(\text{Pa}(G)) - 1$ 
02     do  $G \leftarrow \text{Pa}(G)$ 
03         weise  $\coprod_S(G)$  ein Element der Menge
04              $\{\text{FIND-BASKET}(Z) \in \mathcal{S} \mid Z \in \text{Ch}(G)\}$ 
05             mit kleinster Tiefe zu
06         INSERT-INTO-BASKET(G,  $\coprod_S(G)$ )
07 if  $G \neq \text{Ro}(\mathcal{G})$ 
08     then  $\text{visited-children}(\text{Pa}(G)) \leftarrow \text{visited-children}(\text{Pa}(G)) + 1$ 

```

1. `FIND-BASKET(G)` findet den Basket in welchem G beinhaltet ist. Dazu wird der Knoten in \mathcal{S} ausgegeben, dem dieser Basket zugeordnet ist.
2. `INSERT-INTO-BASKET(G,S)` fügt G in `basket(S)` ein.

7.2.2 Eine Verifikation des LCA-Algorithmus

In diesem Abschnitt zeigen wir anhand der folgenden Lemmata in Theorem 7.1, daß der LCA-Algorithmus das LCA-Problem korrekt löst.

Lemma 7.1

Sei $G \in \mathcal{G}$ in `basket(S)` mit $S \in \mathcal{S}$ enthalten, dann gilt $\coprod_S(G) \subseteq S$.

Beweis. Das Element G ist nur dann in `basket(S)` enthalten, wenn G initial in einen Basket, `basket(I)` mit $I \in \mathcal{S}$, eingefügt wurde. G kann nur in der Prozedur `SPECIES TREE-DFS` von `basket(I)` zu `basket(S)` durch die Operationen `MOVE-BASKET` in Zeile 4 oder `MERGE-BASKET` in Zeile 8 bewegt werden. Dabei werden keine Elemente von `basket(X)` zu `basket(Y)` mit $\text{DEPTH}(X) < \text{DEPTH}(Y)$ bewegt. Daraus folgt: $\text{DEPTH}(I) \geq \text{DEPTH}(S)$ oder in der Mengensystem-Notation $I \subseteq S$.

Initial kann das Element G in `basket(I)` nur durch die Operationen `CREATE-BASKET` in Zeile 10 der Prozedur `SPECIES TREE-DFS` oder die Operation `INSERT-INTO-BASKET` in Zeile 6 der Prozedur `GENETREE-WALK` eingefügt werden. Die Operation `CREATE-BASKET` fügt das Element G initial in `basket($\coprod_S(G)$)` ein, woraus $I = \coprod_S(G)$ folgt. Mit Operation `INSERT-INTO-BASKET` wird das Element G initial in `basket($\coprod_S(G)$)` eingefügt, woraus $I = \coprod_S(G)$ folgt. Also gilt allgemein $I = \coprod_S(G)$.

Somit folgt $\coprod_S(G) \subseteq S$ aus $I = \coprod_S(G)$ und $I \subseteq S$. □

Verw. von: ; Verw. in: Lem. 7.6.

Lemma 7.2

Sei $S \in \mathcal{S}$ und $Z \in \text{Ch}(S)$ das erste Kind von S , in welches die Prozedur SPECIES TREE-DFS von S aus absteigt. $\text{Basket}(S)$ existiert genau dann, wenn der Operationsaufruf $\text{MOVE-BASKET}(Z, S)$ in Zeile 4 nach der Terminierung des Prozeduraufrufes $\text{SPECIES TREE-DFS}(Z)$ ausgeführt worden ist und der Operationsaufruf $\text{MOVE-BASKET}(S, \text{Pa}(S))$ in Zeile 4 oder $\text{MERGE-BASKET}(S, \text{Pa}(S))$ in Zeile 8 nach Terminierung des Prozeduraufrufes $\text{SPECIES TREE-DFS}(S)$ noch nicht erfolgt ist.

Beweis. Nur in der Prozedur SPECIES TREE-DFS kann S ein Basket zugeordnet werden. Diese Zuordnung erfolgt durch die Operationen CREATE-BASKET in Zeile 10 oder MOVE-BASKET in Zeile 4 dieser Prozedur. Da $Z \in \text{Ch}(S)$ gilt, ist S kein Blatt. Die Operation CREATE-BASKET wird nur für Blätter von \mathcal{S} und damit nicht für S ausgeführt. S kann also nur durch die Operation MOVE-BASKET ein Basket zugeordnet werden, welches nur nach der Terminierung des Prozeduraufrufes $\text{SPECIES TREE-DFS}(Z)$ in Zeile 3 der Prozedur SPECIES TREE-DFS geschieht. Das Entfernen von $\text{basket}(S)$ erfolgt damit nur direkt nach der Terminierung des Prozeduraufrufes $\text{SPECIES TREE-DFS}(S)$ durch die Operation MOVE-BASKET in Zeile 4 oder der Operation MERGE-BASKET in Zeile 8 der Prozedur SPECIES TREE-DFS. \square

Verw. von: ; *Verw. in:* Lem. 7.3.

Definition 7.1

Sei L_i ein Blatt von \mathcal{S} , welches gerade von der Prozedur SPECIES TREE-DFS besucht wird, und L_0, \dots, L_{i-1} seien alle Blätter, welche durch die Prozedur SPECIES TREE-DFS schon vorher besucht worden sind. Wir bezeichnen mit $P(L_i)$ den Pfad von $\sqcup(\{L_0, \dots, L_i\})$ zu dem Blatt L_i in \mathcal{S} (der Pfad kann auch durch die Kette $K_{\mathcal{S}}(\sqcup(\{L_0, \dots, L_i\}), L_i)$ beschrieben werden).

Die Bedeutung von $P(L)$ erklärt sich in den beiden folgenden Lemmata.

Lemma 7.3

Sei L ein Blatt von \mathcal{S} , welches gerade von der Prozedur SPECIES TREE-DFS besucht wird. Genau den Elementen des Pfades $P(S)$, die ein Kind besitzen dessen Prozeduraufruf SPECIES TREE-DFS beendet ist deren eigener Prozeduraufruf SPECIES TREE-DFS noch nicht terminiert ist, sind Baskets zugeordnet.

Beweis. Folgt direkt aus Lemma 7.2 \square

Verw. von: Lem. 7.2; *Verw. in:* Lem. 7.6.

Lemma 7.4

Sei L ein Blatt aus S , welches gerade von der Prozedur SPECIES TREE-DFS besucht wird. Weiter sei $S \in S$ und $T \in \text{Ch}(S)$ mit $L \subseteq T$. Gilt $\coprod_S(G) \subset T$ für ein Element G aus $\text{basket}(S)$, so folgt $\coprod_S(G) \notin T_S(T)$.

Beweis. Sei $\coprod_S(G) \subset S$ für ein Element G aus $\text{basket}(S)$. Initial wird G in $\text{basket}(\coprod_S(G))$ eingefügt. Da $\coprod_S(G) \subset S$ gilt und G in $\text{basket}(S)$ enthalten ist, wird G durch die Operationen MERGE-BASKET oder MOVE-BASKET von $\text{basket}(\coprod_S(G))$ dem Inhalt von $\text{basket}(S)$ zugeordnet. Also ist G ein Element von $\text{basket}(U)$ für ein Kind $U \in \text{Ch}(S)$ gewesen und dann durch die Operation MERGE-BASKET oder die Operation MOVE-BASKET dem Inhalt von $\text{basket}(S)$ zugeordnet worden. Diese beiden Operationen können nur nach der Terminierung des Prozeduraufrufes SPECIES TREE-DFS(U) ausgeführt werden.

Da $L \subseteq T$ ist und die Prozedur SPECIES TREE-DFS gerade das Blatt L besucht, ist der Prozeduraufruf SPECIES TREE-DFS(T) noch nicht terminiert. Damit folgt $T \neq U$.

Somit kann der initiale Basket von G , $\text{basket}(\coprod_S(G))$, nicht einem Element aus $T_S(T)$ zugeordnet sein, woraus $\coprod_S(G) \notin T_S(T)$ folgt. \square

Verw. von: ; Verw. in: Lem. 7.6.

Lemma 7.5

Der Algorithmus berechnet $\coprod_S(G)$ für jedes $G \in \mathcal{G}$.

Beweis. Durch vollständige Induktion über S .

I.A.: Sei $G \in \mathcal{G}$ ein Blatt, so ist $\coprod_S(G)$ initial der Wert $\coprod_S(G)$ zugeordnet.

I.V.: Sei $G \in \mathcal{G}$ kein Blatt, und für jedes $Z \in \text{Ch}(G)$ ist der Wert $\coprod_S(Z)$ berechnet worden.

I.S.: Offensichtlich wird jeder Knoten von \mathcal{G} durch den Algorithmus besucht, also auch G . Da G kein Blatt ist, wird dieser Knoten in Zeile 2 der Prozedur GENETREE-WALK besucht. In den folgenden Zeilen 3 bis 5 wird dann $\coprod_S(G)$ berechnet. Dies ist dann möglich, wenn jedes Kind von G in einem Basket enthalten ist, welches wir nun zeigen.

Sei $Z \in \text{Ch}(G)$. Ist Z ein Blatt von S , so ist dieses durch den Prozeduraufruf GENETREE-WALK(Z) besucht worden. Zuvor ist Z initial in $\text{basket}(\coprod_S(Z))$ in Zeile 10 der Prozedur GENETREE-WALK durch die Operation CREATE-BASKET eingefügt worden. Ist Z kein Blatt, so ist nach I.V. $\coprod_S(Z)$ berechnet worden. Diese Berechnung hat in den Zeilen 3 bis 5 der Prozedur GENETREE-WALK stattgefunden. Nach dieser Berechnung ist Z in Zeile 6 initial in $\text{basket}(\coprod_S(Z))$ eingefügt worden.

Da keine Elemente aus Baskets gelöscht werden, sind somit bei der Berechnung von $\coprod_S(G)$ alle Kinder von G in Baskets enthalten. \square

Verw. von: ; Verw. in: Theo. 7.1.

Lemma 7.6

Sei $G \in \mathcal{G}$. Hat der Algorithmus $\coprod_S(G)$ berechnet, so gilt $\coprod_S(G) = \sqcup_S(G)$.

Beweis. Wir zeigen dies durch vollständige Induktion über \mathcal{G} .

I.A.: Durch die Initialisierung gilt für jedes Blatt $L \in \mathcal{G}$ offensichtlich $\coprod_S(L) = \sqcup_S(L)$. Diese Zuordnung ist invariant gegenüber dem Ablauf des Algorithmus. Somit gilt die Aussage für die Blätter von \mathcal{G} .

I.V.: Sei $G \in \mathcal{G}$ ein innerer Knoten, für welchen der Algorithmus den Wert $\coprod_S(G)$ berechnet hat. Es gelte $\sqcup_S(Z) = \coprod_S(Z)$ für jedes Kind $Z \in \text{Ch}(G)$.

I.S.: Wir zeigen zunächst $\sqcup_S(G) \subseteq \coprod_S(G)$. Da G ein innerer Knoten von \mathcal{G} ist, wird $\coprod_S(G)$ in den Zeilen 3 bis 5 der Prozedur `GENETREE-WALK` berechnet. Also ist jedes Kind von G in einem Basket enthalten. Die Berechnung von $\coprod_S(G)$ erfolgt durch einen Prozeduraufruf `GENETREE-WALK($\coprod_S^{-1}(L)$)` für ein Blatt $L \in \mathcal{S}$. Nach Lemma 7.3 sind damit alle Kinder von G in Baskets enthalten, die nur Elementen aus $P(L)$ zugeordnet sind. Also existiert genau ein maximales Element $S \in \mathcal{S}$, welchem ein Basket zugeordnet ist, der ein Element aus $\text{Ch}(G)$ enthält.

Sei $Z \in \text{Ch}(G)$, dann ist Z in $\text{basket}(T)$ für ein $T \in P(L)$ mit $T \subseteq S$ enthalten. Nach Lemma 7.1 gilt $\coprod_S(Z) \subseteq T$. Aus $\coprod_S(Z) \subseteq T$, $T \subseteq S$ und der I.V. folgt $\sqcup_S(Z) \subseteq S$. Also gilt $\sqcup_S(Z) \subseteq S$ für jedes $Z \in \text{Ch}(G)$, woraus $\sqcup_S(G) \subseteq S$ folgt. Da $\coprod_S(G)$ als S berechnet wird, gilt somit $\sqcup_S(G) \subseteq \coprod_S(G)$.

Abschließend zeigen wir $\sqcup_S(G) = \coprod_S(G)$ durch die folgende Fallunterscheidung.

1. $\text{Basket}(S)$ enthält ein $Z \in \text{Ch}(G)$ mit $\coprod_S(Z) = S$: Mit der I.V. folgt, daß $\sqcup_S(Z) = S$ gilt. Da $\coprod_S(G)$ als S berechnet wurde und $\sqcup_S(G) \subseteq \coprod_S(G)$ gilt, folgt $\sqcup_S(G) = S$.
2. $\text{Basket}(S)$ enthält kein $Z \in \text{Ch}(G)$ mit $\coprod_S(Z) = S$: Sei $U \in \text{Ch}(S)$ mit $L \subseteq U$. Wir zeigen, daß ein $Z \in \text{Ch}(G)$ in $\text{basket}(S)$ mit $\sqcup_S(Z) \notin T_S(U)$ existiert. $\text{Basket}(S)$ enthält ein Element $Z \in \text{Ch}(G)$. Nach Lemma 7.1 gilt $\coprod_S(Z) \subseteq S$. Da kein $Z' \in \text{Ch}(G)$ mit $\coprod_S(Z') = S$ existiert, folgt $\coprod_S(Z) \subset S$. Nach Lemma 7.4 folgt damit $\coprod_S(Z) \notin T_S(U)$. Mit der I.V. folgt weiter $\sqcup_S(Z) \notin T_S(U)$.

Sei $Y \in \text{Ch}(G)$ das Kind, welches von dem Prozeduraufruf `GENETREE-WALK($\amalg_S^{-1}(L)$)` besucht wird. Wir zeigen, daß $\sqcup_S(Y) \in T_g(U)$ gilt. Nachdem die Prozedur `GENETREE-WALK` Y in Zeile 2 besucht hat, wird Y in Zeile 6 initial in `basket($\amalg_S(Y)$)` eingefügt. Vor der Berechnung von $\amalg_S(G)$, in dem folgenden Durchlauf der `while`-Schleife, wird Y keinem anderen Basket zugewiesen. Nach Lemma 7.3 ist $\amalg_S(Y) \in P(L)$, womit $\amalg_S(Y) \in T_g(U) \cup S$ folgt. Da kein $Z' \in \text{Ch}(G)$ mit $\amalg_S(Z') = S$ existiert, ist $\amalg_S(Y) \in T_g(U)$. Somit folgt mit der I.V. $\sqcup_S(Y) \in T_g(U)$.

Also gilt $\sqcup_S(G) \subseteq \amalg_S(G)$, es existiert ein $Z \in \text{Ch}(G)$ in `basket(S)` mit $\sqcup_S(Z) \notin T_g(U)$ und ein $Y \in \text{Ch}(G)$ mit $\sqcup_S(Y) \in T_g(U)$. Da $\amalg_S(G) = S$ gilt, folgt hieraus $\sqcup_S(G) = \amalg_S(G)$.

Somit gilt in jedem Fall $\sqcup_S(G) = \amalg_S(G)$. □

Verw. von: Lem. 7.1, 7.3, 7.4; Verw. in: Theo. 7.1.

Theorem 7.1

Der Algorithmus berechnet $\sqcup_S(G)$ für jedes $G \in \mathcal{G}$.

Beweis. Folgt aus Lemma 7.5 und Lemma 7.6 □

Verw. von: Lem. 7.5, 7.6; Verw. in: .

Die Terminierung des Algorithmus ist offensichtlich.

7.2.3 Zeitkomplexität

In diesem Abschnitt zeigen wir die Laufzeit $O(n \alpha(n, n))$ des LCA-Algorithmus. Kostet jede Operation auf Baskets $O(1)$ Zeit, so besitzt der LCA-Algorithmus eine Laufzeit von $O(n)$. Die Laufzeit des Algorithmus ist damit abhängig von dem Zeitaufwand aller Operationen auf Baskets und damit von der Datenstruktur, mit welcher die Baskets dargestellt werden.

Als Datenstruktur für die Baskets verwenden wir die *Basket-Datenstruktur*, welche eine unwesentlich modifizierte Variante der *Disjoint-set-Datenstruktur* (Cormen et. al, 1996: 441) ist. Die Disjoint-set-Datenstruktur verwaltet ein Mengensystem \mathfrak{B} von disjunkten dynamischen Mengen.² Dabei wird jede Menge $\mathcal{B} \in \mathfrak{B}$ eindeutig durch ein Element dieser Menge repräsentiert. Außerdem unterstützt die Disjoint-set-Datenstruktur die folgenden Operationen.

²Mengen in der Informatik können, im Gegensatz zu Mengen in der Mathematik, durch Algorithmen über die Zeit verändert werden. Wir bezeichnen die veränderbaren Mengen der Informatik als dynamische Mengen.

1. MAKE-SET(B) fügt die Menge $\{B\}$ dem Mengensystem \mathfrak{B} hinzu, welches durch die Menge B repräsentiert wird. Dabei ist vor der Anwendung der Operation MAKE-SET(B) das Element B in keiner Menge des Mengensystems \mathfrak{B} enthalten.
2. UNION(C, D). Seien \mathcal{B}_C und \mathcal{B}_D die Mengen in dem Mengensystem \mathfrak{B} , welche durch die Elemente C und D repräsentiert werden. Die Operation UNION(C, D) vereinigt die Mengen \mathcal{B}_C und \mathcal{B}_D miteinander, weist der Menge $\mathcal{B}_C \cup \mathcal{B}_D$ als neuen Repräsentanten ein Element aus $\mathcal{B}_C \cup \mathcal{B}_D$ zu und löscht die Mengen \mathcal{B}_C sowie \mathcal{B}_D aus dem Mengensystem \mathfrak{B} .
3. FIND-SET(E) gibt den Repräsentanten der Menge in \mathfrak{B} zurück, welche das Element E beinhaltet.

Für m MAKE-SET-Operationen (also existieren höchstens $m-1$ UNION-Operationen) und f FIND-SET-Operationen benötigt man nach Cormen et. al (1996:449) eine Laufzeit von $O(m \alpha(m, f))$ (bei Pfadkompression und gewichtetem Union).

Baskets beinhalten paarweise disjunkte Teilmengen der Menge \mathcal{G} und sind Elementen aus \mathcal{S} zugeordnet. Die Disjoint-set-Datenstruktur verwaltet paarweise disjunkte Mengen, welche jeweils durch ein Element der Menge repräsentiert werden. Wählen wir als Repräsentant jeder Menge A der Disjoint-set-Datenstruktur ein Element der Menge \mathcal{S} anstatt eines Elementes von A , so entsprechen die Operationen MAKE-SET, UNION und FIND-SET der Disjoint-set-Datenstruktur jeweils den Operationen CREATE-BASKET, MERGE-BASKET, FIND-BASKET der Basket-Datenstruktur. Somit benötigen alle Operationen CREATE-BASKET, MERGE-BASKET, FIND-BASKET die Zeit $O(m \alpha(m, f))$ für m CREATE-BASKET-Operationen und f FIND-BASKET-Operationen.

Die Operation CREATE-BASKET wird nur bei dem Besuch eines Blattes aus \mathcal{S} durch die Prozedur SpeciesTree-DFS in Zeile 10 aufgerufen, woraus $m = n$ folgt. Die Operation FIND-BASKET wird durch die Prozedur GENETREE-WALK höchstens für jedes Element aus \mathcal{G} aufgerufen, also gilt $f \leq 2n - 1$. Damit benötigen alle CREATE-BASKET-, FIND-BASKET- und MERGE-BASKET-Operationen eine Laufzeit von $O(n \alpha(n, n))$.

Eine Operation MOVE-BASKET(C, D) ordnet der Menge C den neuen Repräsentanten D zu, welches in der Zeit $O(1)$ möglich ist. Der Algorithmus verwendet höchstens $O(n)$ MOVE-BASKET Operationen. Also benötigen alle Operationen auf Baskets eine Zeit $O(n \alpha(n, n))$.

Somit ist die Laufzeit des LCA-Algorithmus $O(n \alpha(n, n))$.

7.3 Eine Reduktion des LCA-Problemes

Die Berechnung aller Duplikationselemente in \mathcal{G} ist durch die Verwendung einer Reduktion auf die Berechnung von kleinsten gemeinsamen Vorfahren im Speziesbaum \mathcal{S} möglich. Dazu zeigen wir zunächst in Abschnitt 7.3.1 eine Linearzeit-Reduktion des LCA-Problemes auf die Berechnung von kleinsten gemeinsamen Vorfahren in \mathcal{S} . In Abschnitt 7.3.2 geben wir dann aus der Literatur bekannte Algorithmen für unterschiedliche Maschinenmodelle zur Berechnung kleinster gemeinsamer Vorfahren in \mathcal{S} an.

7.3.1 Eine Linearzeit-Reduktion

Das LCA-Problem läßt sich durch die beiden folgenden Schritte in Linearzeit auf das Problem der Berechnung von Elementen $\sqcup(Y)$ mit $Y \in \mathcal{S}$ zurückführen (siehe auch Zhang (1997) für binäre Phylogenien).

1. Wir berechnen für jedes $X \in \mathcal{G}$ in postorder das Element $\sqcup(\{\sqcup_{\mathcal{S}}(Z) \in \mathcal{S} \mid Z \in \text{Ch}(X)\})$. Für ein Element $X \in \mathcal{G}$ gilt nach Voraussetzung $\sqcup_{\mathcal{S}}(X) = X$, wenn X ein Blatt ist und sonst $\sqcup_{\mathcal{S}}(X) = \sqcup(\{\sqcup_{\mathcal{S}}(Z) \in \mathcal{S} \mid Z \in \text{Ch}(X)\})$. Damit berechnet der Durchlauf von \mathcal{G} in postorder $\sqcup_{\mathcal{S}}(X)$ für jedes $X \in \mathcal{G}$.
2. Seien Z_1, \dots, Z_l die Kinder von X . Das Element $\sqcup(\{\sqcup_{\mathcal{S}}(Z) \in \mathcal{S} \mid Z \in \text{Ch}(X)\})$ läßt sich durch k_{l-1} wie folgt berechnen:

$$k_1 := \sqcup_{\mathcal{S}}(Z_1) \sqcup \sqcup_{\mathcal{S}}(Z_2) \quad \text{und}$$

$$k_i := k_{i-1} \sqcup \sqcup_{\mathcal{S}}(Z_{i+1}) \quad \text{für jedes } i \in \{2, \dots, l-1\}$$

Damit benötigt man $l-1$ Berechnungen von \sqcup , um $\sqcup_{\mathcal{S}}(X)$ zu berechnen.

Somit benötigt man $O(n)$ Berechnungen von \sqcup , um $\sqcup_{\mathcal{S}}(X)$ für jedes $X \in \mathcal{G}$ zu berechnen.

7.3.2 Berechnungen von \sqcup

Das Problem der Berechnung des kleinsten gemeinsamen Vorfahren \sqcup zweier Elemente in einer Menge von Bäumen ist in den 70er und 80er Jahren intensiv studiert worden (Aho et. al., 1976; Tarjan, 1979; Harel und Tarjan, 1984). Harel und Tarjan (1984:338) formulieren unter anderem das folgende *Off-Line* Problem:

Die Menge der Bäume ist statisch, und die \sqcup -Abfragen sind fest vorgegeben.

Sei n' die Anzahl aller Knoten in den Bäumen und m' die Anzahl der \sqcup -Abfragen. Algorithmen zur Berechnung des Off-line Problemes wurden für unterschiedliche Maschinenmodelle, die *Pointer-Maschine (PM)* (Tarjan, 1979) und die *Random Access Maschine (RAM)* (Aho et al., 1974), entwickelt. Beide Modelle unterscheiden sich in der Darstellung ihres Speichers.

In einer PM besteht der Speicher aus einer Menge von Knoten (welche auch als Records bezeichnet werden). Jeder Knoten beinhaltet eine Menge von *Feldern*. Den Feldern ist jeweils ein Typ wie *pointer*, *integer* oder *real* zugeordnet. Ein Feld eines bestimmten Typs kann einen Wert dieses Typs beinhalten. Z.B. kann ein Feld des Typs *pointer* einen Pointer auf einen Knoten des Speichers der PM beinhalten.

In einer RAM besteht der Speicher aus einem (unendlichen) Array von Wörtern, wobei jedes Wort die Binärdarstellung einer natürlichen Zahl beinhaltet.

Der wesentliche Unterschied zwischen PM und RAM besteht darin, daß bei einer RAM im Gegensatz zu einer PM die Verwendung von Adreßarithmetik möglich ist. Für die Analyse des Zeitaufwands auf einer RAM verwenden Harel und Tarjan das uniforme Kostenmaß und beschränken die Wortlänge auf $O(\log(n'))$ Bits.

Die folgende Tabelle gibt die bisher besten bekannten Laufzeiten für das Off-Line-Problem, jeweils auf einer RAM und einer PM, an.

Zeit auf einer PM	Zeit auf einer RAM
$O(n' + m' \alpha(m' + n', n'))$	$O(n')$
Aho, Hopcroft und Ullman (1976)	Harel and Tarjan (1984)

Tabelle 7.1: Die schnellsten PM und RAM Algorithmen für das Off-Line Problem

Betrachten wir unser ursprüngliches Problem, das Element $\sqcup_{\mathcal{G}}(X)$ für jedes $X \in \mathcal{G}$ zu berechnen. Bei fest vorgegebenem Genbaum \mathcal{G} sind alle \sqcup -Anfragen fest vorgegeben, womit ein Off-Line-Problem zu lösen ist. Wir haben gezeigt, daß $O(n)$ \sqcup -Anfragen benötigt werden, um $\sqcup_{\mathcal{G}}(X)$ für jedes $X \in \mathcal{G}$ zu berechnen. Damit kann nach Tabelle 7.1 auf einer PM $\sqcup_{\mathcal{G}}(X)$ für jedes $X \in \mathcal{G}$ in der Zeit $O(n \alpha(n, n))$ und auf einer RAM in Zeit $O(n)$ berechnet werden.

Der LCA-Algorithmus verwendet keine Adreßarithmetik und arbeitet direkt auf Knoten, womit dieser Algorithmus auf einer PM lauffähig ist. Asymptotisch

laufen der LCA-Algorithmus und der modifizierte Algorithmus von Aho, Hopcroft und Ullman beide in der Zeit $O(n \alpha(n, n))$. Jedoch ist der \sqcup_S -Algorithmus im Gegensatz zu dem modifizierten Algorithmus von Aho, Hopcroft und Ullman direkt auf das \sqcup_S -Problem zugeschnitten und benötigt eine, nicht asymptotisch betrachtete, geringere Laufzeit als der modifizierte Algorithmus.

Eine untere Schranke ist dem Autor sowohl für das \sqcup_S -Problem, als auch für das Off-Line Problem nicht bekannt. Für die von dem LCA-Algorithmus verwendete Basket-Datenstruktur können wir jedoch, für m' MERGE-BASKET und f' FIND-BASKET Operationen, auf einer PM die untere Schranke $\Omega(m' \alpha(m', f'))$ direkt von der unteren Schranke (Tarjan, 1979) für die Disjoint-set-Datenstruktur ableiten.

7.4 Die Berechnung des Maßes M_1

Um das Maß M_1 aus Abschnitt 6 für \mathcal{S} und \mathcal{G} zu berechnen, verwenden wir Lemma 6.1. Danach ist $M_1(\mathcal{G}, \mathcal{S}) = \sum_{X \in \mathcal{G}, Z \in \text{Ch}(X)} l(X, Z)$, wobei l für jedes $X \in \mathcal{G}$ und $Z \in \text{Ch}(X)$ wie folgt definiert ist.

$$l(X, Z) := \begin{cases} |K_S^-(\sqcup_S(X), \sqcup_S(Z))| + 1 & X \in \text{DupK} \wedge \sqcup_S(Z) \neq \sqcup_S(X) \\ |K_S^-(\sqcup_S(X), \sqcup_S(Z))| & X \notin \text{DupK} \\ 0 & \text{sonst.} \end{cases}$$

Um die reduzierten Kettenlängen zu berechnen, ordnen wir zunächst in der Initialisierung jedem Element $U' \in \mathcal{S}$ seine Höhe $\text{depth}(U')$ zu. Für jedes $X \in \mathcal{G}$ und $Z \in \text{Ch}(Z)$ wird der Wert $l'(X, Z) \leftarrow \text{depth}(Z) - \text{depth}(X)$ berechnet, welches der reduzierten Kettenlänge $|K_S^-(\sqcup_S(X), \sqcup_S(Z))| + 1$ entspricht. Falls X kein Duplikationselement ist, also $\text{Duplikationselement}(X) = \text{FALSE}$ gilt, decrementieren wir den Wert von $l'(X, Z)$ um eins. Somit gilt $l'(X, Z) = l(X, Z)$.

Die Tiefe der Knoten ist in Zeit $O(n)$ berechenbar, und $l'(X)$ berechnet sich in der Zeit $O(1)$, womit $l'(X, Z)$ sich für alle $X \in \mathcal{G}$ und $Z \in \text{Ch}(X)$ in Zeit $O(n)$ berechnet. Durch die Addition aller Werte für l' erhalten wir damit in Zeit $O(n)$ den Wert für $M_1(\mathcal{G}, \mathcal{S})$.

Die Duplikationselemente sowie das Maß $M_1(\mathcal{G}, \mathcal{S})$ sind somit in Zeit $O(n)$ berechenbar, wenn $\sqcup_S(X)$ für alle $X \in \mathcal{G}$ bekannt ist.

Kapitel 8

Paraloge Evolution in der Farbwahrnehmung

In diesem Kapitel zeigen wir die Erforschung der Evolution der Farbwahrnehmung von Wirbeltieren durch die Biologen Kawamura und Yokoyama anhand des konstruktiven Reconciled Tree.

Kawamura und Yokoyama untersuchen die Evolution der Farbspektren, in welchen Organismen Licht wahrnehmen. Wirbeltiere verwenden dazu visuelle Pigmente, mit denen sie Licht unterschiedlicher Wellenlängen wahrnehmen und damit in Nervensignale umwandeln. Ein visuelles Pigment besteht aus einem Chromophor (ein Vitamin A Derivat, A_1 oder A_2), welcher an das membranständige Enzym Opsin gebunden ist. Bei der Umwandlung von Lichtenergie in ein Nervensignal wird zunächst die Lichtenergie in eine lichtinduzierte Konformationsveränderung des Chromophores umgewandelt, worauf dann das Opsin aktiviert wird. Die Wellenlänge λ_{\max} , bei welcher ein visuelles Pigment die maximale Lichtenergie absorbiert (Absorptionsspektrum), wird durch das Opsin und den Chromophor bestimmt.

Beim Menschen besitzen z.B. alle visuellen Pigmente den gleichen Chromophor (Vitamin A_1). Unterschiedliche Absorptionsspektren λ_{\max} werden nur durch unterschiedliche Opsine bestimmt. Es gibt aber Amphibien und Fische, welche ihr Farbspektrum schnell an die Umwelt anpassen mußten und somit nicht die notwendige Zeit besaßen, die Aminosäuresequenz ihres Opsins zu verändern. Die Evolution sicherte das Überleben dieser Spezies, indem sie das Absorptionsspektrum λ_{\max} eines visuellen Pigments durch Verwendung eines anderen Chromophores (Vitamin A_2 oder A_1) festlegte¹.

¹Es gibt Fische und Vögel, bei denen λ_{\max} durch einen Lichtfilter, bestehend aus einem kleinen farbigen Öltröpfchen festgelegt ist (Okano et al., 1992; Wang et al., 1992). Eine allgemeine

Die Protein-Familie der Opsine ist unterteilt in die vier Opsin-Familien: RH1 (Rhodopsine), RH2 (rhodopsinähnlich mit unterschiedlichen Absorptionsspektren λ_{\max}), SWS (kurze Wellenlängen (blau) λ_{\max}) und LWS/MWS (lange und mittlere Wellenlängen (grün, rot) λ_{\max}).

Die Evolution der Opsin-Familie RH2 ist aufgrund ihrer biologischen Vielfalt an unterschiedlichen Absorptionsspektren λ_{\max} besonders interessant. Kawamura und Yokoyama untersuchen für diese Unterfamilie, wann und wie sich die unterschiedlichen Absorptionsspektren λ_{\max} der visuellen Pigmente entwickelt haben, also wann in der Evolution eine Verschiebung des Absorptionsspektrums λ_{\max} stattgefunden hat und durch welche Mechanismen (Aminosäuresequenz des Opsins, Chromophor, ...) dies erreicht wurde.

1995 sequenzierten Kawamura und Yokoyama (1995) das RH2-Opsin OPSB_ANOCA (P51471) des Tag-Kamelions *Anolis Carolinensis* mit unbekannter Funktion. Obwohl die Funktion von OPSB_ANOCA unbekannt war, gab diese ein vollkommen neues Bild der Evolution der RH2-Opsine. Denn aufgrund der Sequenzinformation von OPSB_ANOCA postulierten Kawamura und Yokoyama eine Genduplikation in der Evolution der RH2-Opsine, welche die Existenz bisher unbekannter RH2-Opsine vermuten ließ und eine neue Hypothese zu der Verschiebung der Wellenlänge λ_{\max} in der Evolution der RH2-Opsine etablierte.

Wir werden in diesem Kapitel den konstruktiven Reconciled Tree zu dem RH2-Opsin OPSB_ANOCA konstruieren. Durch den Reconciled Tree werden wir die Annahme, daß OPSB_ANOCA von einer Genduplikation abstammt, untermauern. Aber der Reconciled Tree zeigt auch die Entwicklung dieser Genduplikation, an welcher wir die Analysen und Resultate von Kawamura und Yokoyama nachvollziehen werden.

In Abschnitt 8.1 werden wir den Reconciled Tree zu dem RH2-Opsin OPSB_ANOCA konstruieren und diesen dann in Abschnitt 8.2 analysieren.

Dies ist ein Anwendungsbeispiel dafür, daß durch die Postulierung von Genduplikationen neue Mechanismen der Evolution aufgedeckt werden können. In diesem Fall haben an der Evolution der Opsine interessierte Biologen gezielt für ihre Forschung das Gen OPSB_ANOCA sequenziert, um damit die Evo-

Beschreibung der Funktion von Opsinen ist in Stryer (1979) zu finden.

lution der RH2-Opsine weiter zu entschlüsseln. Doch in Sequenzdatenbanken existieren Unmengen an Sequenzen, deren Bezug zu Genduplikation unbekannt ist und mit denen man Einsichten in die Mechanismen der Evolution gewinnt. Mit der hier beschriebenen Methode können ganze Sequenzdatenbanken automatisiert auf Duplikationsereignisse hin untersucht werden, welche die Kenntnis über die Evolution der Funktionalität von Genen revolutionieren kann.

8.1 Konstruktion des Reconciled Tree

Um in Abschnitt 8.1.3 den Reconciled Tree \mathcal{RK} für das RH2-Opsin OPSB_ANOCA zu rekonstruieren, benötigen wir einen Genbaum \mathcal{G} und einen Speziesbaum \mathcal{S} . Den Genbaum rekonstruieren wir in Abschnitt 8.1.2 für Opsine, welche homolog zu dem RH2-Opsin OPSB_ANOCA sind (dies sollten alle bekannten RH2-Opsine sein). Den Speziesbaum \mathcal{S} erhalten wir für die Spezies dieser RH2-Opsine aus der Taxa-Datenbank des NCBI. Wie wir die zu dem RH2-Opsin OPSB_ANOCA homologen Sequenzen erhalten, ist in dem folgenden Abschnitt gezeigt.

8.1.1 Homologe Sequenzen zu OPSB_ANOCA

Um zu dem RH2-Opsin OPSB_ANOCA homologe Proteine zu finden, durchsuchen wir zunächst die Sequenzdatenbank SWISSPROT (Bairoch und Boeckmann, 1994) mit BLAST (Altschul et. al, 1990) nach Sequenzen, die dem Protein OPSB_ANOCA ähnlich sind. Die Tabelle 8.1 zeigt das Ergebnis einer solchen Suche. Je geringer der *p-value* Wert eines Proteines ist, desto wahrscheinlicher ist es, daß die zugehörige Sequenz dem Protein OPSB_ANOCA ähnlich ist.

AC	ID	CC	p-value
P51471	OPSB_ANOCA	?-SENSITIVES RH2-OPSIN	0.0
P28683	OPSG_CHICK	GRÜN-SENSITIVES RH2-OPSIN	e-174
P35357	OPSB_GECGE	BLAU-SENSITIVES RH2-OPSIN	e-157
P32311	OPSG_CARAU	GRÜN-SENSITIVES RH2-OPSIN	e-147
P22671	OPSD_LAMJA	RHODOPSIN	e-145
P32312	OPSH_CARAU	GRÜN-SENSITIVES RH2-OPSIN	e-145
P51474	OPSI_ASTFA	GRÜN-SENSITIVES RH2-OPSIN	e-140
P56515	OPSD_BUFMA	RHODOPSIN	e-138
P56514	OPSD_BUFBU	RHODOPSIN	e-137
P22328	OPSD_CHICK	RHODOPSIN	e-136
P29403	OPSD_XENLA	RHODOPSIN	e-136
P51470	OPSD_RANCA	RHODOPSIN	e-136
P41591	OPSD_ANOCA	RHODOPSIN	e-136
P52202	OPSD_ALLMI	RHODOPSIN	e-136
P31355	OPSD_RANPI	RHODOPSIN	e-135
P32308	OPSD_CANFA	RHODOPSIN	e-135
P56516	OPSD_RANTE	RHODOPSIN	e-135
P79863	OPSD_RAJER	RHODOPSIN	e-135
P49912	OPSD_RABIT	RHODOPSIN	e-134
Q90245	OPSD_AMBTI	RHODOPSIN	e-134
O18766	OPSD_PIG	RHODOPSIN	e-133
Q28886	OPSD_MACFA	RHODOPSIN	e-133
P08100	OPSD_HUMAN	RHODOPSIN	e-133

Tabelle 8.1: BLAST-Suche mit OPSD_ANOCA (P51471)

Alle gefundenen Sequenzen sind Opsine von Wirbeltieren. Ein multiples Alignment mit CLUSTAL W (Thompson et al., 1994) der Sequenzen aus Tabelle 8.1 (ein Teil dieses Alignments ist in Abbildung 8.1 gezeigt) zeigt die starke Homologie des Opsins OPSD_ANOCA zu den RH2-Opsinen, welche in der folgenden Tabelle 8.2 aufgeführt sind.

AC	ID	Funktion und Spezies
P28683	OPSG_CHICK	grün-sensitives RH2-Opsin des Huhns
P32311	OPSG_CARAU	grün-sensitives RH2-Opsin des Goldfisches
P32312	OPSH_CARAU	grün-sensitives RH2-Opsin des Goldfisches
P35357	OPSB_GECGE	blau-sensitives RH2-Opsin des Nacht-Eidechsen-Geckos
P51474	OPSL_ASTFA	grün-sensitives RH2-Opsin des blinden Höhlenfisches

Tabelle 8.2: Opsine der RH2-Familie

Somit werden wir den Genbaum \mathcal{G} für die folgende Menge von RH2-Opsinen rekonstruieren.

$$M := \{\text{OPSD_ANOCA}, \text{OPSG_CHICK}, \text{OPSG_CARAU}, \\ \text{OPSH_CARAU}, \text{OPSB_GECGE}, \text{OPSL_ASTFA}\}$$

8.1.2 Rekonstruktion von \mathcal{G} und \mathcal{S}

Um den Genbaum für die RH2-Opsine der Menge M aus dem vorigen Abschnitt zu rekonstruieren, verwenden wir die Methode *Neighbour-Joining* (Saitou und Nei, 1987). Diese Methode rekonstruiert den Genbaum für die Proteine der Menge M anhand einer Distanz-Matrix, welche paarweise die evolutionären Abstände zwischen allen Sequenzen der Menge M angibt. Ergebnis dieses Verfahrens ist ein auf dieser Distanz-Matrix basierender ungewurzelter Genbaum mit den Blättern der Menge M .

Eine allgemeine Vorgehensweise, um den Genbaum für die RH2-Opsine zu wurzeln, ist es, den RH2-Opsinen eine Menge von Opsinen einer anderen Opsinfamilie hinzuzufügen. Diese Menge wird allgemein als *Outgroup* bezeichnet. Mit dieser Menge rekonstruiert man dann den Genbaum, in welchem die Verzweigung zwischen der Outgroup und den RH2-Opsinen dann die Wurzel für den Teilbaum der RH2-Opsine angibt.

Als Outgroup verwenden wir die RH1-Opsine der Menge

$$M' := \{\text{OPSD_HUMAN}, \text{OPSD_LAMJA}, \text{OPSD_ANOCA}\}$$

und rekonstruieren den ungewurzelten Genbaum für die Opsine der Menge $M \cup M'$, welchen wir dann mit \mathcal{G} bezeichnen.

Um die Distanz-Matrix für die Sequenzen der Menge $M \cup M'$ zu erhalten, erzeugen wir mit CLUSTAL W das in Abbildung 8.1 gezeigte *multiple Alignment* für die Menge $M \cup M'$.

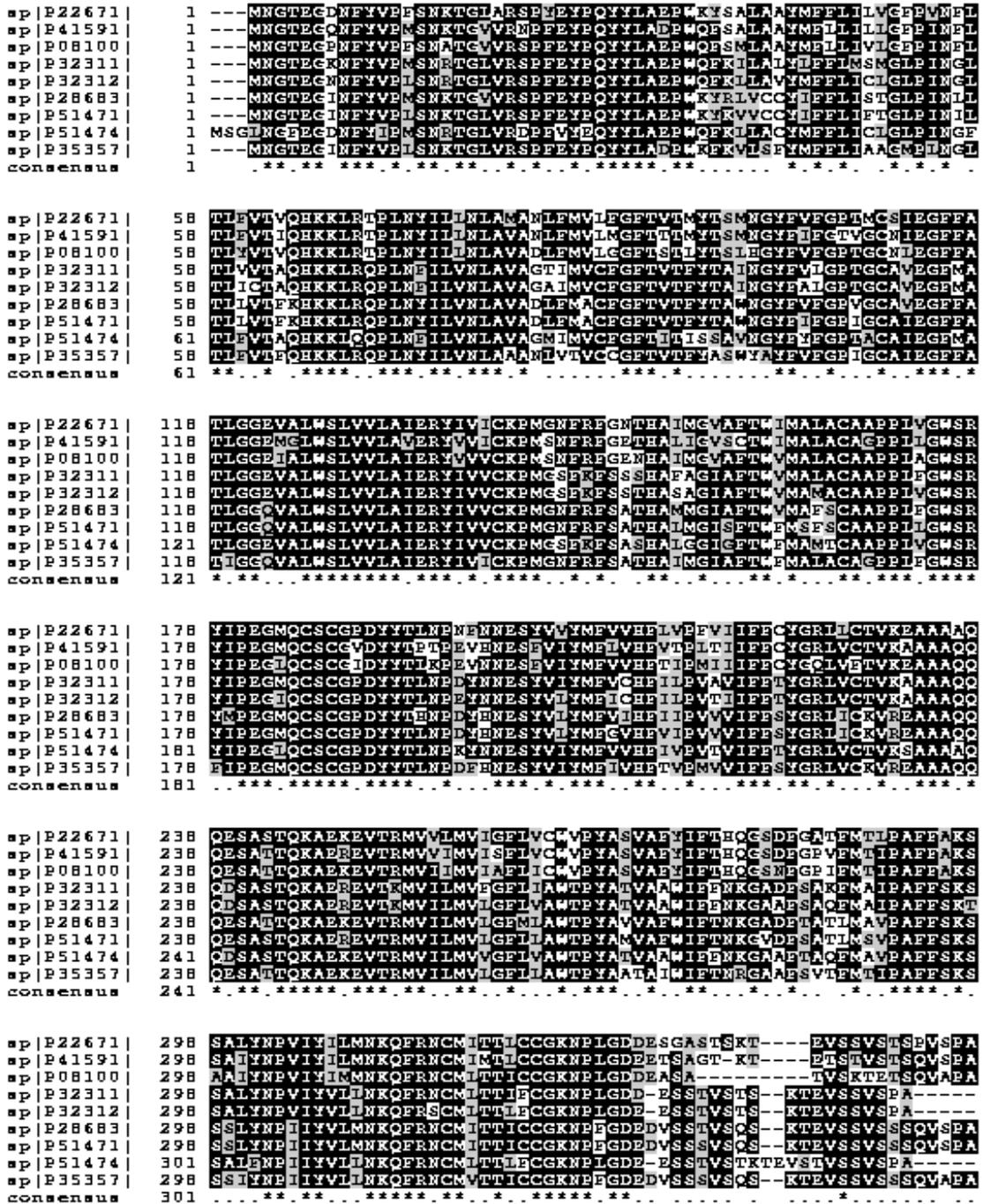


Abbildung 8.1: Multiples Alignment

Hieraus ergibt sich dann die Distanz-Matrix, aus welcher der in Abbildung 8.2 gezeigte ungewurzelte Genbaum für die Menge $M \cup M'$ durch die Methode *Neighbour-Joining* rekonstruiert wurde.

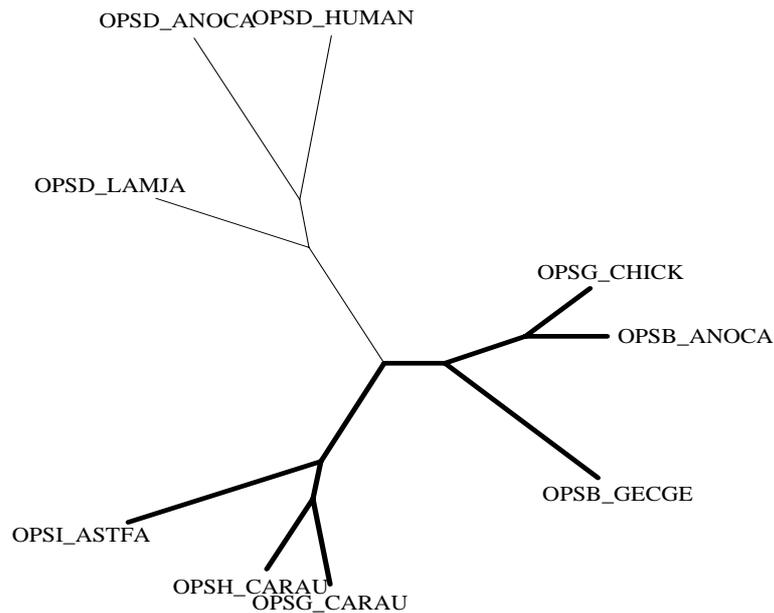


Abbildung 8.2: Ungewurzelter Genbaum

An diesem ungewurzelten Baum ist deutlich die evolutionäre Verzweigung zwischen den RH1- und RH2-Opsinen zu erkennen. Diese verwenden wir nun, um hieraus den gewurzelten Genbaum der RH2-Opsine \mathcal{G} aus Abbildung 8.3 zu erhalten. Den Genbaum \mathcal{G} haben wir dabei in seine Darstellung als n-Clustersistem überführt. Um die Mengen des n-Clustersistemes einfacher darzustellen, haben wir Mengen, die mehr als ein Element beinhalten, durch die Buchstaben a, b, c, d gekennzeichnet.

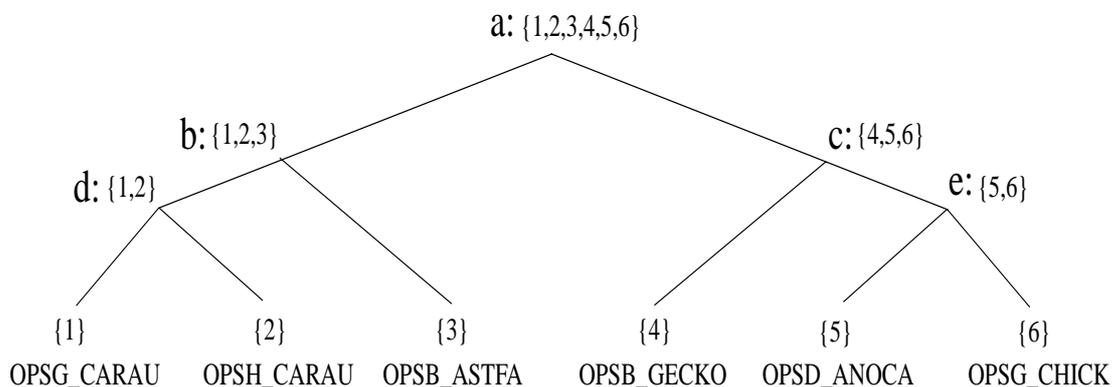


Abbildung 8.3: Gewurzelter Genbaum \mathcal{G}

Den Speziesbaum \mathcal{S} der Gene, welcher in Abbildung 8.4 gezeigt ist, erhalten wir für die Spezies der Proteinsequenzen aus der Taxa-Datenbank des NCBI. Diese enthält eine nicht aufgelöste Verzweigung, eine Multifurkation, an der Wurzel. Selbst bei wenig Sequenzen, wie in unserem Fall, enthält die Taxa-Datenbank häufig Multifurkationen, so daß die Berücksichtigung von Multifurkationen bei der Vorhersage von Genduplikationen notwendig ist.

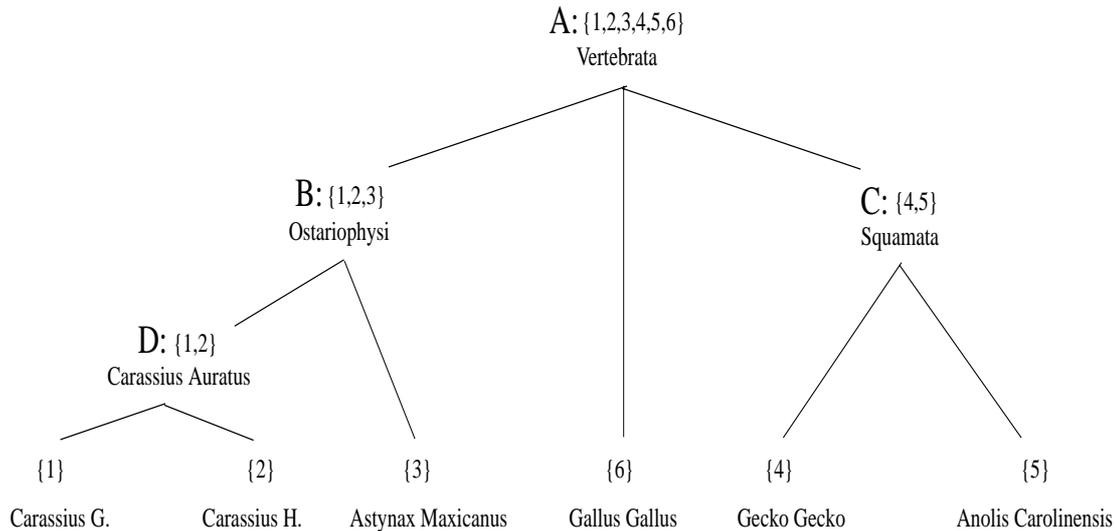


Abbildung 8.4: Speziesbaum \mathcal{S}

Der Speziesbaum ist in seiner Darstellung als n -Clusterelementsystem gezeigt. Zur Vereinfachung sind, wie bei dem Genbaum \mathcal{G} , die Mengen mit Buchstaben gekennzeichnet. Die lateinischen Namen der heute existenten Spezies entsprechen folgenden umgangssprachlichen Bezeichnungen:

Carassius Auratus - Goldfisch, Astyanax fasciatus - blinder Höhlenfisch,
Gallus Gallus - Huhn, Gecko Gecko - Gecko, Anolis Carolinensis -
Tag-Kamelion.

Die RH2-Opsine OPSG_CARAU und OPH_CARAU gehören beide der Spezies Carassius Auratus (Goldfisch) an. Um diese bei der Rekonstruktion des Reconciled Tree zu berücksichtigen, haben wir, wie am Ende von Abschnitt 2.2.2 beschrieben, diesen jeweils eine künstliche Spezies *CarassiusG* und *CarassiusH* zugeordnet.

8.1.3 Bestimmung des Reconciled Tree für \mathcal{G} und \mathcal{S}

In dem vorigen Abschnitt haben wir die n -Clusterelementsysteme \mathcal{G} und \mathcal{S} rekonstruiert. Hieraus konstruieren wir nun den Reconciled Tree \mathcal{RK} von dem Genbaum

\mathcal{G} über dem Speziesbaum \mathcal{S} . Wir zeigen zur Anschauung vollständig die Konstruktion von \mathcal{RK} , obwohl der Aufwand für dieses kleine Beispiel schon die Anwendung eines Programms unter Verwendung eines LCA-Algorithmus aus Kapitel 7 rechtfertigt. Im folgenden sind noch einmal die n -Clusterelemente \mathcal{G} und \mathcal{S} angegeben.

$$\begin{aligned}\mathcal{G} &= \{a, b, c, d, e, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\} \\ \mathcal{S} &= \{A, B, C, D, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}\end{aligned}$$

Nach Definition 5.2 ist der konstruktive Reconciled Tree \mathcal{RK} wie folgt definiert:

$$\begin{aligned}\text{BG} &= \{(X, X, S) \in \mathcal{G} \times \mathcal{G} \times \mathcal{S} \mid S = \bigsqcup_S(X)\} \\ \text{UG} &= \{(X, Z, S) \in \mathcal{G} \times \mathcal{G} \cup \{?\} \times \mathcal{S} \mid \\ &\quad X \in \overline{\text{DupK}} \wedge (Z \in \text{Ch}(X) \wedge S \in e_S(Z) \\ &\quad \vee Z = ? \wedge S \in \tilde{E}_S(X)) \\ &\quad \vee X \in \text{DupK} \wedge S \in T_S(\bigsqcup_S(X)) \setminus T_S(\bigsqcup_S(Z))\}\} \\ \mathcal{RK} &= \text{BG} \cup \text{UG}\end{aligned}$$

Wir fassen in Abschnitt 8.1.3.1 zunächst die zur Konstruktion von \mathcal{RK} notwendigen Definitionen zusammen. In Abschnitt 8.1.3.2 bestimmen wir dann die Menge \mathcal{RK} und geben in Abschnitt 8.1.3.3 den Reconciled Tree als das Hasse-Diagramm der auf \mathcal{RK} durch \subseteq gegebenen Baumordnung an.

8.1.3.1 Notwendige Definitionen für \mathcal{RK}

Wir zeigen zunächst die Definition des Duplikationselements in \mathcal{G} und folgend die Definitionen der lokalen und unbestimmten Entwicklungsumgebung in \mathcal{S} .

Duplikationselement in \mathcal{G} :

Nach Definition 5.1 ist die Menge der Duplikationselemente in \mathcal{G} wie folgt definiert:

$$\text{DupK} = \{X \in \mathcal{G} \mid \neg P_{\text{NoDup}}(X)\}.$$

Dabei ist das Prädikat $P_{\text{NoDup}}(X)$ nach Definition 4.15 genau dann wahr, wenn der folgende Ausdruck gilt:

$$\bigsqcup_S(Z_i) \subset \bigsqcup_S(X) \text{ und } \bigsqcup_S(Z_i) \sqcup \bigsqcup_S(Z_j) = \bigsqcup_S(X).$$

Lokale und unbestimmte Entwicklungsumgebung in \mathcal{S} :

Für ein Element $X \in \mathcal{G}$ und $Z \in \text{Ch}(X)$ ist nach Definition 4.13 (Seite 72) die lokale Entwicklungsumgebung $e_s(Z)$ und nach Definition 4.14 (Seite 72) die unbestimmte Entwicklungsumgebung $\tilde{E}_s(X)$ in \mathcal{S} wie folgt definiert:

$$\begin{aligned} e_s(Z) &= \{Y' \in \mathcal{S} \mid \exists C' \in \text{Ch}(\bigsqcup_s(\text{Pa}(Z))) : \\ &\quad \bigsqcup_s(Z) \subset C' \Rightarrow Y' \in T_s(C') \setminus T_s(\bigsqcup_s(Z))\} \\ \tilde{E}_s(X) &= E_s(X) \setminus (\{\bigsqcup_s(X)\} \cup \bigcup_{Z \in \text{Ch}(X)} e_s(Z)) \end{aligned}$$

8.1.3.2 Bestimmung der Menge \mathcal{RK}

Wir bestimmen \mathcal{RK} , welche die Vereinigung der bekannten Gene BG und der unbekannt Genen UG ist, schrittweise. Die bekannten Gene in \mathcal{RK} sind alle Dreitupel der folgenden Menge.

$$\begin{aligned} BG &= \{(a, a, A), (b, b, B), (c, c, A), (d, d, D), (e, e, A), \\ &\quad (1, 1, 1), (2, 2, 2), (3, 3, 3), (4, 4, 4), (5, 5, 5), (6, 6, 6)\} \end{aligned}$$

Die unbekannt Genen UG werden aus der Vereinigung der beiden folgenden Mengen gebildet:

$$\begin{aligned} UG_1 &:= \{(X, Z, S) \in \mathcal{G} \times \mathcal{G} \cup \{?\} \times \mathcal{S} \mid X \in \overline{\text{DupK}} \wedge (Z \in \text{Ch}(X) \wedge S \in e_s(Z) \\ &\quad \vee Z = ? \wedge S \in \tilde{E}_s(X))\} \\ UG_2 &:= \{(X, Z, S) \in \mathcal{G} \times \mathcal{G} \cup \times \mathcal{S} \mid X \in \text{DupK} \wedge S \in T_s(\bigsqcup_s(X)) \setminus T_s(\bigsqcup_s(Z))\} \end{aligned}$$

Für die Mengen UG_1 und UG_2 sind die Duplikationselemente aus \mathcal{G} sowie die lokalen und unbestimmten Entwicklungsumgebungen von \mathcal{S} zu bestimmen. Hierzu benötigen wir für jedes Element $G \in \mathcal{G}$ dessen kleinsten gemeinsamen Vorfahren $\bigsqcup_s(G)$ in \mathcal{S} , welcher der Host-Spezies von G entspricht. Es gilt:

$$\begin{aligned} \bigsqcup_s(a) &= A, \bigsqcup_s(b) = B, \bigsqcup_s(c) = A, \bigsqcup_s(d) = D, \bigsqcup_s(e) = A \\ \text{und } \bigsqcup_s(y) &= y \quad \text{für jedes } y \in \{1, \dots, 6\}. \end{aligned}$$

Bestimmung der Duplikationselemente:

Ein Element X ist genau dann ein Duplikationselement in \mathcal{G} , wenn das Prädikat $P_{\text{NoDup}}(X)$ nicht gilt. Die Aussage $\bigsqcup_s(Z_i) \sqcup \bigsqcup_s(Z_j) = \bigsqcup_s(X)$ ist für alle $X \in \mathcal{G}$ und ungleiche $Z_i, Z_j \in \text{Ch}(X)$ wahr. Die Aussage $\bigsqcup_s(Z_i) \subset \bigsqcup_s(X)$ ist genau für die Elemente $\mathcal{G} \setminus \{a, c\}$ wahr. Also sind die Elemente $a, c \in \text{DupK}$ Duplikationselemente und die Elemente $d, e, \{1\}, \dots, \{6\} \in \overline{\text{DupK}}$ keine Duplikationselemente in \mathcal{G} .

Bestimmung der Menge UG_1 :

Die Menge UG_1 setzt sich somit aus den Dreitupeln der Formen $(b, ,)$, $(d, ,)$ und $(e, ,)$ zusammen, da $b, d, e \in \overline{\text{DupK}}$ gilt. Für die Form $(e, ,)$ existieren die folgenden Dreitupel:

$(e, ?,)$: $(e, ?, B), (e, ?, D), (e, ?, \{1\}), (e, ?, \{2\}), (e, ?, \{3\})$. Der dritte Parameter ergibt sich jeweils aus den Elementen der unbestimmten Entwicklungsumgebung $\tilde{E}_S(e)$ (siehe Abb. 8.5).

$(e, 5,)$: $(e, 5, C), (e, 5, \{4\})$. Der dritte Parameter ergibt sich jeweils aus den Elementen der lokalen Entwicklungsumgebung $e_S(5)$ (siehe Abb. 8.5).

$(e, 6,)$: Die lokale Entwicklungsumgebung $e_S(6)$ ist die leere Menge, (siehe Abb. 8.5).

Für die Formen $(b, ,)$ und $(d, ,)$ kann der dritte Parameter, wie in Abbildung 8.5 gezeigt ist, nicht belegt werden. Somit existieren für diese Formen keine Dreitupel, und es gilt:

$$UG_1 = \{(e, ?, B), (e, ?, D), (e, ?, \{1\}), (e, ?, \{2\}), (e, ?, \{3\}), (e, 5, C), (e, 5, \{4\})\}$$

$$b: \tilde{E}_S(b) = e_S(d) = e_S(3) = \emptyset$$

$$e: \tilde{E}_S(e) = \{B, D, \{1\}, \{2\}, \{3\}\}$$

$$e_S(5) = \{C, \{4\}\}$$

$$d: \tilde{E}_S(d) = e_S(1) = e_S(2) = \emptyset$$

$$e_S(6) = \emptyset$$

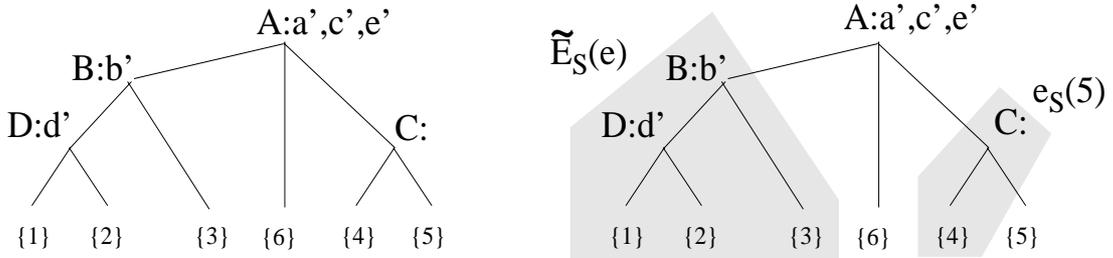


Abbildung 8.5: Zur Konstruktion von UG_1 (dabei gilt $y' = \sqcup_S(y)$)

Bestimmung der Menge UG_2 :

Die Menge UG_2 setzt sich aus den Dreitupeln der Formen $(a, ,)$ und $(c, ,)$ zusammen, da $a, c \in \text{DupK}$ gilt. Für die Form $(a, ,)$ existieren die folgenden Dreitupel:

(a, c) : Da $T_S(\sqcup_S(a)) \setminus T_S(\sqcup_S(c)) = \emptyset$ gilt, existieren keine Dreitupel dieser Form (siehe Abb. 8.6).

(a, d): (a, d, A), (a, d, C), (a, d, {6}), (a, d, {4}), (a, d, {5}). Der dritte Parameter ergibt sich jeweils aus den Elementen der Menge $T_S(\bigsqcup_S(a)) \setminus T_S(\bigsqcup_S(d))$ (siehe Abb. 8.6).

Für die Form (c, ,) existieren die folgenden Dreitupel:

(c, e): Da $T_S(\bigsqcup_S(c)) \setminus T_S(\bigsqcup_S(e)) = \emptyset$ gilt, existieren keine Dreitupel dieser Form (siehe Abb. 8.6).

(c, {4}): (c, {4}, A), (c, {4}, B), (c, {4}, D), (c, {4}, {1}), (c, {4}, {2}), (c, {4}, {1}), (c, {4}, {3}), (c, {6}, {1}), (c, {4}, {5}). Der dritte Parameter ergibt sich jeweils aus den Elementen der Menge $T_S(\bigsqcup_S(c)) \setminus T_S(\bigsqcup_S(\{4\}))$ (siehe Abb. 8.6).

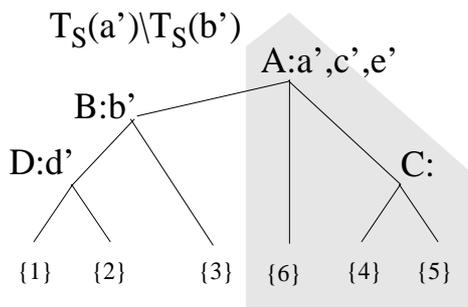
Somit gilt:

$$UG_2 = \{(c, \{4\}, A), (c, \{4\}, B), (c, \{4\}, D), \\ (c, \{4\}, \{1\}), (c, \{4\}, \{2\}), (c, \{4\}, \{1\}), (c, \{4\}, \{3\}), (c, \{6\}, \{1\}), (c, \{4\}, \{5\})\}$$

$$a: T_S(a') \setminus T_S(c') = \emptyset$$

$$T_S(a') \setminus T_S(b') = \{A, C, \{6\}, \{5\}, \{4\}\}$$

$$T_S(a') \setminus T_S(b')$$



$$c: T_S(c') \setminus T_S(e') = \emptyset$$

$$T_S(c') \setminus T_S(\{4\}) = \{A, B, C, D, \\ \{1\}, \{2\}, \{3\}, \{6\}, \{5\}\}$$

$$T_S(c') \setminus T_S(\{4\})$$

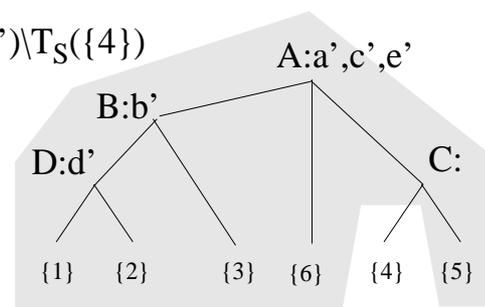


Abbildung 8.6: Zur Konstruktion von UG_2 (dabei gilt $y' = \bigsqcup_S(y)$)

8.1.3.3 Hasse-Diagramm der Baumordnung \subseteq auf \mathcal{RK}

Nachdem wir alle Elemente der Menge \mathcal{RK} kennen, können wir deren Elemente nach der in Definition 5.4 (Seite 99) formulierten Baumordnung anordnen. Abbildung 8.7 zeigt das Hasse-Diagramm dieser Baumordnung. Schraffierte Bereiche kennzeichnen die unbekanntene Gene von \mathcal{RK} . Die Kennzeichnungen eines schraffierten Bereiches geben die Host-Spezies aller Gene dieses Bereiches und damit die Menge der dritten Parameter der Dreitupel des Bereiches an. Alle

unterstrichenen Dreitupel sind bekannte Gene. Damit wird die Zusammensetzung von RK aus der Topologie von \mathcal{G} und den Subtopologien von \mathcal{S} ersichtlich. Die fettgedruckten Kanten zeigen die Einbettung des Genbaumes \mathcal{G} in den Reconciled Tree RK. Nicht gestrichelte fettgedruckte Kanten verbinden dabei alle bekannte Gene des in RK eingebetteten Genbaumes \mathcal{G} . Die Dreitupel (a, a, A) und (c, c, A) sind Duplikationselemente in \mathcal{RK} , da diese die Duplikationselemente a und c in \mathcal{G} darstellen.

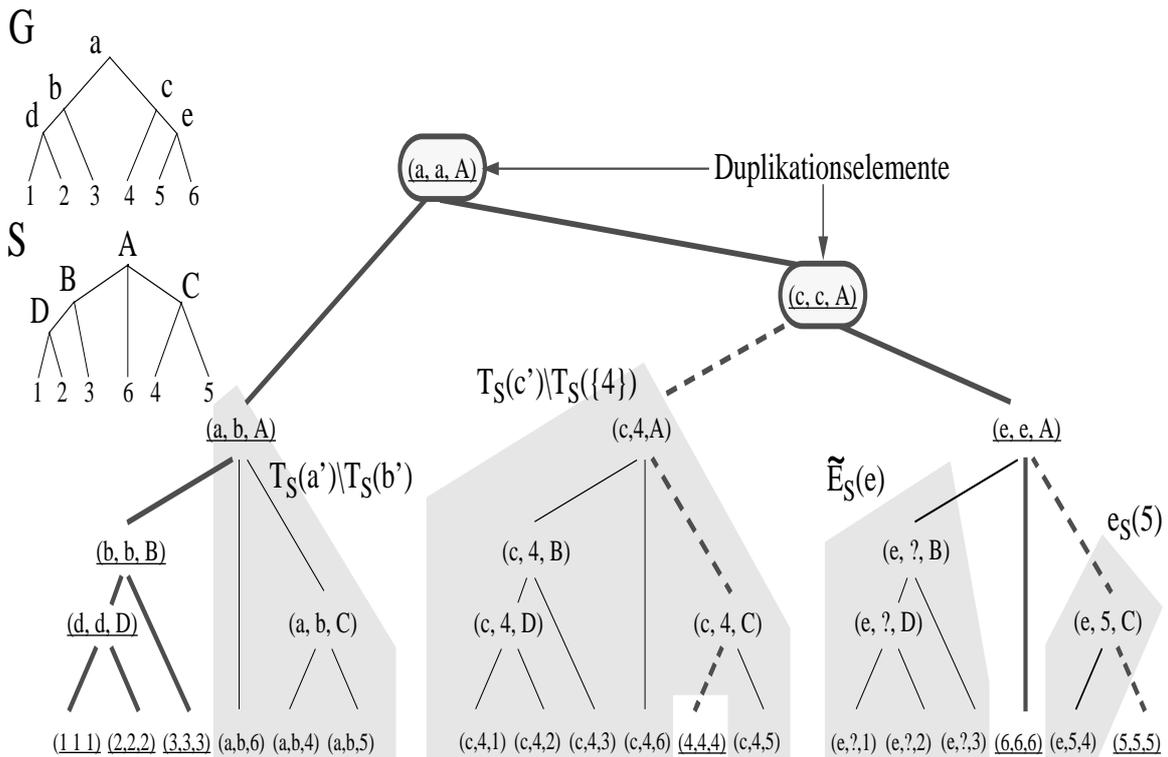


Abbildung 8.7: Der Reconciled Tree \mathcal{RK}

8.2 Analyse des Reconciled Tree

Der Reconciled Tree in Abbildung 8.8 zeigt die Evolution der RH2-Opsine. Wie auch von den Biologen Kawamura und Yokoyama postuliert, sagt der Reconciled Tree Genduplikationen in der Familie der RH2-Opsine voraus. Durch die Genduplikation an den Genen (a, a, A) und (c, c, A) erhalten wir somit eine neue, paraloge Vorstellung von der Evolution der RH2-Opsine.

Um hieraus Aussagen über die Entwicklung der Absorptionsspektren von RH2-Opsinen abzuleiten, betrachten wir zu diesen Opsinen alle bisher bekannten Absorptionsspektren. In Tabelle 8.3 haben wir dazu die bekannten Absorptionsspektren.

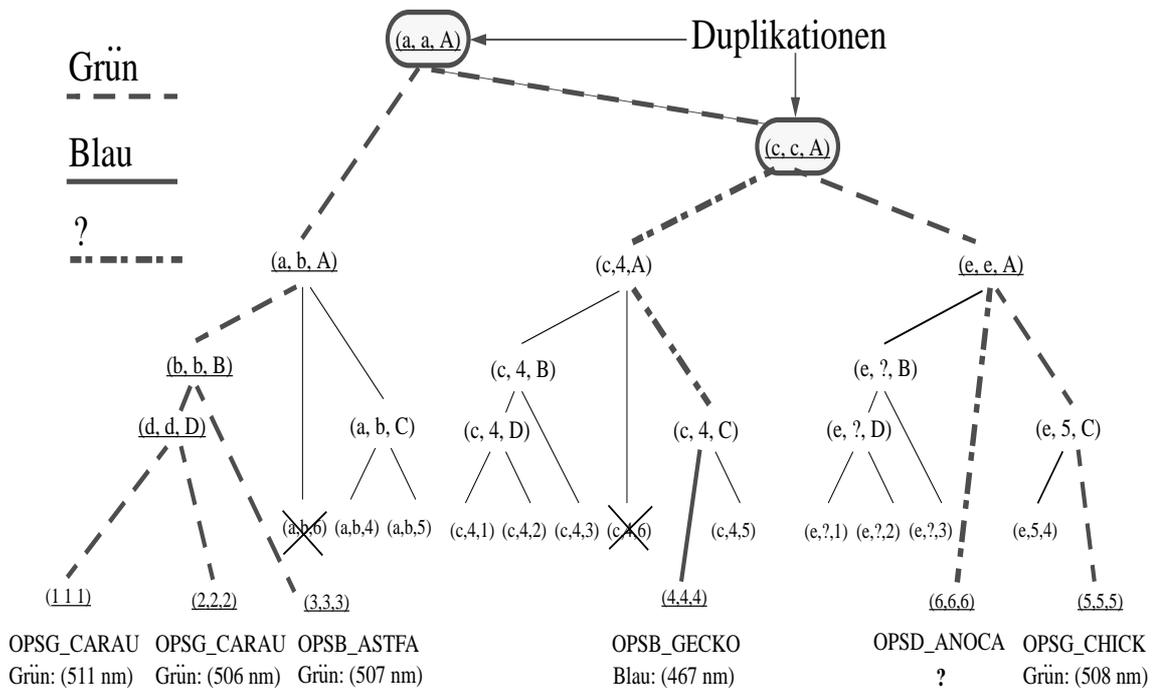


Abbildung 8.8: Eine Hypothese am Reconciled Tree RK

spektren von RH2-Opsinen zusammengefaßt und den jeweiligen RH2-Opsinen des Reconciled Tree in Abbildung 8.8 zugeordnet² (Referenzen zu den Absorptionsspektren der RH2-Opsine sind dem Artikel von Kawamura und Yokoyama (1995) zu entnehmen.)

ID	λ_{\max} (nm ⁻¹)	Farbspektrum	Chromophor
OPSB_GECGE	467	blau	A1
OPSH_CARAU	506	grün	A1
OPSB_ASTFA	507	grün	A1
OPSG_CHICK	508	grün	A1
OPSG_CARAU	511	grün	A1
OPSD_ANOCA	?	?	A2

Tabelle 8.3: Funktion von RH2-Opsinen

Mit den bekannten grünen Absorptionsspektren der RH2-Opsine können wir nun zwei Hypothesen zu deren Entwicklung zeigen. Betrachten wir dazu die

²Bis heute hat man bei Wirbeltieren nur in visuellen Pigmenten von *Alonis Carolinensis* das Vitamin A₂ als Chromophor gefunden. Bei allen anderen visuellen Pigmenten von Wirbeltieren hat man bisher ausschließlich das Vitamin A₁ als Chromophor gefunden (Provencio et al., 1992).

Entwicklung des grünen Absorptionsspektrums der RH2-Opsine OP SH_CARAU, OPSG_CARAU, OPSB_ASTFA von Fischen und dem RH2-Opsin OPSG_CHICK des Huhns. Zu der Entwicklung des grünen Absorptionsspektrums der Fische und des Huhns gibt es die folgenden Hypothesen:

Divergente Evolution: Das Protein des Gens (α, α, A) besitzt ein grünes Absorptionsspektrum, dieses hat sich dann zu den grünen Absorptionsspektren der RH2-Opsine OP SH_CARAU, OPSG_CARAU, OPSB_ASTFA und OPSG_CHICK entwickelt.

Konvergente Evolution: Das grüne Absorptionsspektrum der RH2-Opsine OP SH_CARAU, OPSG_CARAU, OPSB_ASTFA der Fische und das grüne Absorptionsspektrum des Proteines OPSG_CHICK vom Huhn haben sich unabhängig voneinander entwickelt.

Die Wahrscheinlichkeit ist jedoch gering, daß sich die grüne Absorptionsspektren der Fische und des Huhns über einen langen Zeitraum unabhängig voneinander und in vollkommen unterschiedlichen Lebensbereichen entwickelt haben. So vertreten auch Kawamura und Yokoyama die Hypothese, daß sich die grünen Absorptionsspektren der RH2-Opsine OP SH_CARAU, OPSG_CARAU, OPSB_ASTFA und OPSG_CHICK divergent entwickelt haben.

Nach dieser Hypothese der divergenten Evolution des grünen Absorptionsspektrums hat sich das Absorptionsspektrum des Proteines OPSB_GECGE durch eine Blauverschiebung aus einem grünen Absorptionsspektrum entwickelt. Das blaue Absorptionsspektrum des RH2-Opsins OPSB_GECGE hat sich demnach aus einem der RH2-Opsine (c, c, A), ($c, 4, 6$) oder ($c, 4, C$) entwickelt.

Die Sequenzierung der paralogen RH2-Opsine ($c, 4, 1$), ($c, 4, 2$), ($c, 4, 3$), ($c, 4, 6$) oder ($c, 4, 5$) könnte hierzu eine genauere Aussage ermöglichen. Kawamura und Yokoyama haben dazu das Genom von *Anolis Carolinensis* nach RH2-Opsinen, welche zu dem RH2-Opsin OPSD_ANOCA paralog sind, durchsucht. Ihre Analyse hat jedoch ergeben, daß es mit großer Wahrscheinlichkeit keine solchen paralogen RH2-Opsine gibt. Für die anderen RH2-Opsine ist diese Analyse noch nicht möglich, da die hierzu notwendige Genom-Information nicht vorhanden ist.

Gehen wir davon aus, daß die zu dem RH2-Opsin ($6, 6, 6$) paralogen RH2-Opsine ($3, b, 6$) und ($c, 4, 6$) nicht existieren. Damit existierten möglicherweise auch schon Vorgänger der RH2-Opsine ($3, b, 6$) und ($c, 4, 6$) nicht. Auf die frühere Existenz der direkten Vorgänger (α, b, A) und ($c, 4, A$) dieser RH2-Opsine (und

damit auf die frühere Existenz aller weiteren Vorgänger) gibt es jedoch deutliche Hinweise. Denn das RH2-Opsin (a, b, A) besitzt die heute existenten Nachfolger (1, 1, 1), (2, 2, 2), (3, 3, 3) und das RH2-Opsin (c, 4, A) den heute existenten Nachfolger (4, 4, 4).

Durch das neu sequenzierte Protein OPSD_ANOCA konnten die Biologen Kawamura und Yokoyama und wir mit Hilfe des Reconciled Tree Genduplikationen vorhersagen. Dabei ist uns das Absorptionsspektrum dieses RH2-Opsins immer noch unbekannt. Üblicherweise sagt man dem RH2-Opsin OPSD_ANOCA das Absorptionsspektrum eines stark homologen RH2-Opsins voraus, welche in diesem Fall das grüne Absorptionsspektrum des RH2-Opsin OPSG_CHICK vom Huhn ist.

Aufgrund biologischer Experimente haben Kawamura und Yokoyama jedoch starke Hinweise darauf erhalten, daß das RH2-Opsin OPSD_ANOCA ein blaues Absorptionsspektrum besitzt. Danach haben sich die blauen Absorptionsspektren der RH2-Opsine OPSD_ANOCA und OPSB_GECGE unabhängig voneinander, also konvergent, entwickelt. Kawamura und Yokoyama haben hierzu ein Experiment vorgeschlagen, mit welchem sie in kürze diese Hypothese überprüfen wollen.

Kapitel 9

Schluß: Zusammenfassung und Ausblick

Genduplikationen sind ein wesentlicher Mechanismus der molekularen Evolution, mit deren Kenntnis wichtige biologische Fragestellungen beantwortet werden können (siehe Kapitel 1). In dieser Arbeit haben wir gezeigt, wie Genduplikationen und deren Entwicklung methodisch vorhersagbar sind. Dazu geben wir in Abschnitt 9.1 einen abschließenden Überblick. Neuartige Anwendungen, welche durch diese Arbeit ermöglicht werden, zeigen wir in Abschnitt 9.2.

9.1 Zusammenfassung

In dieser Arbeit haben wir Methoden und Maße zu der Vorhersage von Genduplikationen und deren Entwicklung in der Evolution aufgezeigt. Diese basieren auf dem Modell des Reconciled Tree, welches der Biologe Page 1994 semiformal eingeführt hat. Um unsere Methoden und Maße von diesem Modell abzuleiten, haben wir zunächst einen Reconciled Tree formal definiert. Mit dieser Definition haben wir dann Strukturaussagen zu einem Reconciled Tree gezeigt, auf denen die Resultate dieser Arbeit im wesentlichen beruhen.

Die Strukturaussagen zeigen einerseits die Topologie eines Reconciled Tree zu einem Gen- und Speziesbaum, sowie deren Eindeutigkeit. Andererseits wird durch das Duplikations-Theorem gezeigt, daß die Genduplikationen eines Reconciled Tree alleine durch den Gen- und Speziesbaum berechnet werden können, die diesen definieren.

Mit dem Duplikations-Theorem konnten wir erstmalig die Korrektheit des Page-Algorithmus zeigen. Hypothesen und mögliche aufwendige Experimente von Biologen, die aufgrund der Genduplikationen des Programmes COMPONENT gemacht wurden, sind somit abgesichert.

Basierend auf der uns bekannten eindeutigen Topologie eines Reconciled Tree haben wir zu einem Genbaum \mathcal{G} und Speziesbaum \mathcal{S} den konstruktiven Reconciled Tree definiert. Dieser besitzt dieselbe Topologie wie ein Reconciled Tree \mathcal{R} zu \mathcal{G} über \mathcal{S} und ist damit, für die Vorhersage von Genduplikationen und deren Entwicklung, äquivalent zu \mathcal{R} . Der konstruktive Reconciled Tree ist im Gegensatz zu einem Reconciled Tree biologisch interpretierbar, formal einfach zu handhaben und konstruktiv definiert.

Aus dem konstruktiven Reconciled Tree haben wir dann den Mixed-Tree abgeleitet, welcher die lokale Vorhersage und Analyse von Genduplikationen zuläßt.

Zu dem konstruktiven Reconciled Tree haben wir Maße zu der Vorhersagequalität von Genduplikationen formuliert. Wir haben gezeigt, daß jedes uns aus der Literatur bekannte Maß zu einem von uns definierten Maß äquivalent ist. In der Literatur äquivalente, jedoch unterschiedlich definierte Maße konnten wir hierdurch aufzeigen. Insbesondere haben wir die Mirkin-Muchnik-Smith-Vermutung gezeigt, womit die Äquivalenz zweier, nach völlig unterschiedlichen biologischen Konzepten definierten Maße, bewiesen ist.

Letztendlich sind damit alle uns aus der Literatur bekannten Maße durch die schrittweise Verfeinerung eines, von uns formulierten, biologischen Konzeptes darstellbar. Hieraus leiten wir eine Validierung des von uns definierten biologischen Konzeptes ab.

Mit dem Duplikations-Theorem haben wir einen Algorithmus entwickelt, welcher Genduplikationen und alle uns aus der Literatur bekannten Maße für einen Genbaum \mathcal{G} und einen Speziesbaum \mathcal{S} in einer Zeit $O(n \alpha(n, n))$ mit $n := |\mathcal{G}|$ berechnet. Der Algorithmus ist für dieses Problem maßgeschneidert und dadurch auch in der praktischen Anwendung schnell.

Durch das Duplikations-Theorem konnten wir eine Linearzeit-Reduktion auf bekannte Probleme zeigen, mit welchen sich Genduplikationen auch in einer Zeit $O(n \alpha(n, n))$ berechnen lassen. Diese sind jedoch in der praktischen Anwendung langsamer als der maßgeschneiderte Algorithmus.

Kennen wir die Genduplikationen, so ist der konstruktive Reconciled Tree linear in der Anzahl seiner Elemente, direkt nach seiner konstruktiven Definition berechenbar.

Abschließend haben wir Genduplikationen und deren Entwicklung durch den konstruktiven Reconciled Tree in der Familie der RH2-Opsine vorhergesagt und analysiert. Unsere Vorhersage untermauert die von den Biologen Kawamura

und Yokoyama postulierte Genduplikation in der Familie der RH2-Opsine.

9.2 Ausblick

Mit den Algorithmen und Maßen dieser Arbeit kann nun ein Programm erstellt werden, mit welchem Genduplikationen und deren Entwicklung schneller vorhergesagt und genauer analysiert werden können, als dies durch das Programm COMPONENT möglich ist. Durch den – in dieser Arbeit erstmals eingeführten – Mixed-Tree kann der Biologe dann zusätzlich die vorhergesagten Genduplikationen lokal analysieren. Ein solches Programm sollte dem Biologen *Online* über das WWW kostenlos verfügbar gemacht werden.

Genduplikationen wollen wir jedoch nicht nur aufgrund von durch Biologen vorgegebenen Gen- und Speziesbäumen vorhersagen. Obwohl Biologen gewisse Genfamilien auf Genduplikationen hin untersucht haben, fehlt es bis heute an einer systematischen Analyse aller zur Zeit bekannten Sequenzdaten. In Zukunft sollen durch die in dieser Arbeit vorgestellten Algorithmen und Maße die Genbäume aller Genfamilien der Datenbank SwissProt auf Genduplikationen hin untersucht werden. Dazu soll jeweils der konstruktive Reconciled Tree, welcher die Entwicklung von vorhergesagten Genduplikationen zeigt, berechnet werden.

Aufbauend auf dieser Analyse soll dann ein Stammbaum der untersuchten Spezies aufgestellt werden, der ein einheitliches Modell für die untersuchten Genbäume aller Genfamilien darstellt. Ein solcher Speziesstammbaum ist dann evolutionsbiologisch von großem Wert und erlaubt auch eine Verfeinerung der vorangegangenen Analyse.

Das Ergebnis einer solchen Analyse ist eine umfassende Datenbank von Stammbäumen von Genfamilien. Neben den Genduplikationen sollen in dieser Datenbank auch bisher unbekannte Gene, die sich aus den Genduplikationen entwickelt haben, enthalten sein. Dieser Datensatz ist für die Genomanalyse im speziellen und die molekularbiologische Forschung im allgemeinen von großer Bedeutung. Er erlaubt feinere Unterscheidungen in der Funktionsvorhersage, als dies bisher – basierend auf dem einfachen Homologiebegriff – möglich war. Die generierte Information hilft z.B. bei der Planung der in Kapitel 1 erwähnten *knock-out*-Experimente, die ein wesentliches Instrument zur funktionalen Analyse darstellen. Schließlich eröffnet ein solcher Datensatz die Möglichkeit, gezielt nach Genen zu suchen, deren Existenz im konstruktiven Reconciled Tree

vorausgesagt wurde. Eine solche Suche zeigt das Anwendungsbeispiel in Kapitel 8.

Für die hier vorgestellten Anwendungen stellt diese Arbeit alle benötigten und wesentlichen theoretischen Resultate bereit. Es wird nun darum gehen, diese Anwendungen in die Praxis umzusetzen, um damit dem Biologen ein effektives Hilfsmittel zu der Aufdeckung neuer Mechanismen der Evolution bereitzustellen.

Allgemeine Definitionen

Seien M und N endliche Mengen.

Definition 9.1 (Eigenschaften von Relationen)

Eine Relation R auf M heißt

- reflexiv, wenn $\forall m \in M : m R m.$
- symmetrisch, wenn $\forall m, n \in M : m R n \Rightarrow n R m.$
- antisymmetrisch, wenn $\forall m, n \in M : m R n \wedge n R m \Rightarrow m = n.$
- transitiv, wenn $\forall m, n, o \in M : m R n \wedge n R o \Rightarrow m R o.$
- alternativ, wenn $\forall m, n \in M : m R n \vee n R m.$

Definition 9.2 (Äquivalenzrelation, Ordnung, totale Ordnung)

Eine Relation R in einer Menge M heißt

Äquivalenzrelation, wenn sie reflexiv, symmetrisch und transitiv ist.

Ordnung, wenn sie reflexiv, antisym. und transitiv ist.

Totale Ordnung, wenn sie eine Ordnung und alternativ ist.

(Früher wurde eine Ordnung als Halbordnung oder teilweise Ordnung bezeichnet. Die totale Ordnung wird auch als einfache oder lineare Ordnung bezeichnet (Halmos, 1976).)

Definition 9.3 (Ketten)

Sei \sqsubseteq eine Ordnung auf M und $m, n \in M$ mit $n \sqsubseteq m$, dann bezeichnen wir

$K_M(m, n) := \{k \in M \mid n \sqsubseteq k \sqsubseteq m\}$, als die Kette von n nach m in M .

$K_M^-(m, n) := \{k \in M \mid n \sqsubset k \sqsubset m\}$, als die reduzierte Kette von n nach m in M .

Definition 9.4 (Vergleichbar)

Sei \sqsubseteq eine Ordnung auf M . Zwei Elemente $m, n \in M$ heißen vergleichbar in M , wenn $m \sqsubseteq n$ oder $n \sqsubseteq m$ gilt. Abkürzend schreiben wir $m \diamond n$, wenn m und n miteinander vergleichbar sind.

Definition 9.5 (Obere und untere Schranken)

Sei \sqsubseteq eine Ordnung auf M , $K \subseteq M$ und $s \in K$.

s heißt obere Schranke von K in M
 $:\iff \forall m \in K : m \sqsubseteq s$.

s heißt kleinste obere Schranke von K in M
 $:\iff \forall m \in M : m \text{ ist obere Schranke von } K \Rightarrow s \sqsubseteq m$.

Wir bezeichnen mit $\bigsqcup K$ die kleinste obere Schranke von K in M . Für $K = \{m, n\}$ bezeichnen wir mit $m \sqcup n$ die kleinste obere Schranke der beiden Elemente m und n in M .

Dual definieren sich untere Schranke und größte untere Schranke. Die größte untere Schranke bezeichnen wir mit \sqcap .

Definition 9.6 (Maxima, Minima)

Sei \sqsubseteq eine Ordnung auf M .

$s \in M$ heißt Maximum von M $:\iff \forall m \in M : m \sqsubseteq s$
 $s \in M$ heißt maximales Element von M $:\iff \forall m \in M : m \diamond s \Rightarrow m \sqsubseteq s$

Dual definieren sich Minimum und minimales Element.

Definition 9.7 (Ordnungshomomorphismen)

Sei \sqsubseteq_M eine Ordnung auf M und \sqsubseteq_N eine Ordnung auf N .

$\text{Ohom}(M, N) := \{f : M \rightarrow N \mid \forall x, y \in M : x \sqsubseteq_M y \Rightarrow f(x) \sqsubseteq_N f(y)\}$
 $\text{Omon}(M, N) := \{f : M \rightarrow N \mid f \in \text{Ohom}(M, N) \wedge f \text{ ist injektiv}\}$
 $\text{Oepi}(M, N) := \{f : M \rightarrow N \mid f \in \text{Ohom}(M, N) \wedge f \text{ ist surjektiv}\}$
 $\text{Oiso}(M, N) := \{f : M \rightarrow N \mid f \in \text{Ohom}(M, N) \wedge f \text{ ist bijektiv}\}$

Wir bezeichnen ein Element aus $\text{Ohom}(M, N)$, $\text{Omon}(M, N)$, $\text{Oepi}(M, N)$ oder $\text{Oiso}(M, N)$ jeweils als Ordnungshomomorphismus, Ordnungsepimorphismus oder Ordnungsisomorphismus von M nach N .

Definition 9.8 (Einbettungshomomorphismen)

Sei \sqsubseteq_M eine Ordnung auf M und \sqsubseteq_N eine Ordnung auf N (die kleinste obere Schranke bezieht sich im folgenden jeweils auf die für die Menge angegebene Ordnung).

$\text{Ehom}(M, N) := \{f : M \rightarrow N \mid \forall x, y \in M : f(x \sqcup y) = f(x) \sqcup f(y)\}$
 $\text{Emon}(M, N) := \{f : M \rightarrow N \mid f \in \text{Ehom}(M, N) \wedge f \text{ ist injektiv}\}$
 $\text{Eepi}(M, N) := \{f : M \rightarrow N \mid f \in \text{Ehom}(M, N) \wedge f \text{ ist surjektiv}\}$
 $\text{Eiso}(M, N) := \{f : M \rightarrow N \mid f \in \text{Ehom}(M, N) \wedge f \text{ ist bijektiv}\}$

Wir bezeichnen ein Element aus $E_{\text{hom}}(M, N)$, $E_{\text{mon}}(M, N)$, $E_{\text{epi}}(M, N)$ oder $E_{\text{iso}}(M, N)$ jeweils als Einbettungshomomorphismus, Einbettungsepimorphismus oder Einbettungsisomorphismus von M nach N .

Definition 9.9 (Minimale Funktion)

Sei \sqsubseteq_N eine Ordnung auf N und F eine Menge von Funktionen, welche von M nach N abbilden. Wir bezeichnen eine Funktion $f \in F$ als minimal in F , wenn $\forall x \in M : \forall f' \in F : f(x) \sqsubseteq_N f'(x)$.

Technische Hilfslemmata

Die Hilfslemmata zeigen formal, daß die in Abschnitt 4.4.2 auf den Seiten 74 und 82 gemachten Konstruktionen von \mathcal{R}' sowie \mathcal{R}'' Reconciled Trees ergeben. Sie beschreiben damit formal die Ersetzungen aus Abbildung 4.10 und 4.11. Beide Konstruktionen ersetzen zuerst einen Teilbaum $T_{\mathcal{R}}(Y'')$ in \mathcal{R} durch die Ableitung $T_{\mathcal{S}}(B(Y'')) \times p$ ohne Duplikationselemente von $T_{\mathcal{S}}(B(Y''))$ für ein zu $\text{Ro}(\mathcal{R})$ tdisjunktes $p \in \mathbb{N}$. In dem folgenden Lemma zeigen wir, daß dies eine "erlaubte" Ersetzung ist, d.h. daß die Parameter der Ersetzung ihrem Gültigkeitsbereich entsprechen.

Lemma 9.2 zeigt dann formal die Konstruktion von \mathcal{R}'' aus Abschnitt 4.4.3.1, und Lemma 4.4.3.3 zeigt formal die Konstruktion von \mathcal{R}' aus Abschnitt 4.4.3.3.

Lemma 9.1

Sei $Y'' \in \mathcal{R}$ und $p \in \mathbb{N}$ tdisjunkt zu \mathcal{R} , dann ist $\mathcal{R}' := \text{Repl}(\mathcal{R}, T_{\mathcal{S}}(B(Y'')) \times p, Y'')$ eine Ersetzung.

Beweis. Wir zeigen die Bedingungen der Parameter:

1. \mathcal{R} ist ein m -Clustersystem: nach Voraussetzung ist \mathcal{R} ein Reconciled Tree und damit auch ein m -Clustersystem.
2. $T_{\mathcal{S}}(B(Y'')) \times p$ ist ein m -Clustersystem: Da nach Voraussetzung $Y'' \in \mathcal{R}$, gilt $B(Y'') \in \mathcal{S}$, womit $T_{\mathcal{S}}(B(Y'')) \times p$ existiert. Theorem 4.3 besagt, daß $T_{\mathcal{S}}(B(Y'')) \times p$ ein m -Clustersystem ist.
3. $\text{Ro}(T_{\mathcal{S}}(B(Y'')) \times p) \cap \text{Ro}(\mathcal{R}) = \emptyset$: gilt wegen der Wahl von p .
4. $B(\text{Ro}(T_{\mathcal{S}}(B(Y'')) \times p)) = B(Y'')$: $B(\text{Ro}(T_{\mathcal{S}}(B(Y'')) \times p)) = B(B(Y'') \times \{p\}) = B(Y'')$.
5. $Y'' \in \mathcal{R}$: folgt direkt aus der vorausgesetzten Tdisjunktheit von p .

Somit sind die Bedingungen an die Parameter von \mathcal{R}' erfüllt, und \mathcal{R}' ist eine Ersetzung. \square

Verw. von: Theo. 4.3; Verw. in: Lem. 9.2, 9.3.

Von einer Ersetzung \mathcal{R}^0 ausgehend, ersetzen wir nun entsprechend unterschiedlicher Umwandlungen einen oder alle Teilbäume $T_{\mathcal{R}}(Z'')$ mit $Z'' \in \varphi(\text{Ch}(X))$.

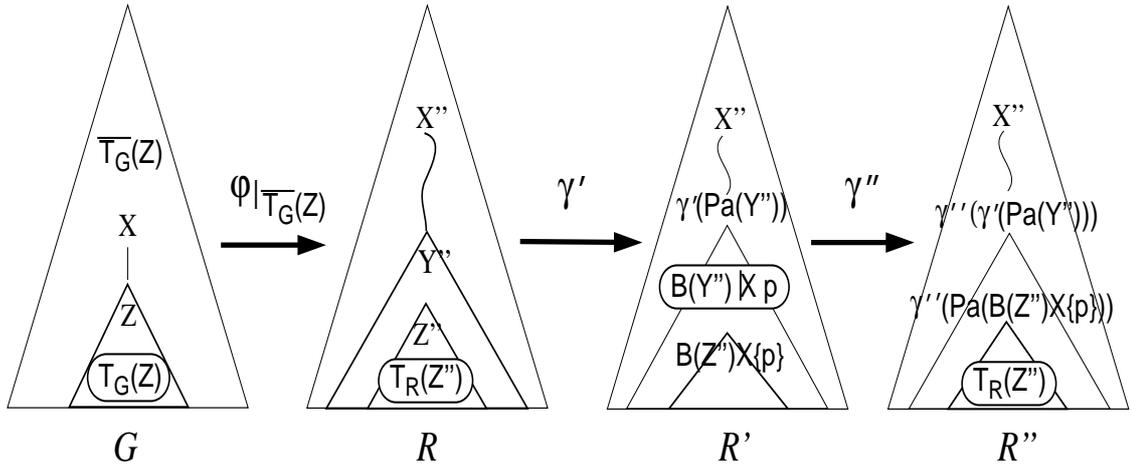


Abbildung 9.1: Die Ersetzung $\text{Repl}(\mathcal{M}, \mathcal{N}, I)$

Lemma 9.2 Seien $X \in \mathcal{G}$ und $Y'' \in \mathcal{R}$, sodas genau ein $Z \in \text{Ch}(X)$ mit $\varphi(Z) \subseteq Y''$ existiert. Weiter sei $\mathcal{R}' := \text{Repl}(\mathcal{R}, T_s(B(Y'')) \times p, Y'')$, für ein zu $\text{Ro}(\mathcal{R})$ tdisjunktes $p \in \mathbb{N}$. Dann ist

$$\mathcal{R}'' := \text{Repl}(\mathcal{R}', T_{\mathcal{R}}(\varphi(Z)), B(\varphi(Z)) \times \{p\}) \in \text{RC}(\mathcal{G}, \mathcal{S})$$

Beweis.

Zur Vereinfachung definieren wir $Z'' := \varphi(Z)$ und $X'' := \varphi(X)$.

Aussage 1 \mathcal{R}'' ist eine Ersetzung.

Beweis. Wir zeigen die Bedingungen für die einzelnen Parameter:

1. \mathcal{R}' ist ein m -Clustersystem: Lemma 9.1 besagt: \mathcal{R}' ist eine Ersetzung. Damit folgt aus Theorem 4.1 die Aussage.
2. $T_{\mathcal{R}}(Z'')$ ist ein m -Clustersystem: nach Voraussetzung ist \mathcal{R} ein Reconciled Tree und damit ein m -Clustersystem. Da jeder Teilbaum eines m -Clustersystemes ein m -Clustersystem ist, gilt dies auch für $T_{\mathcal{R}}(Z'')$.

3. $\text{Ro}(T_{\mathcal{R}}(Z'')) \cap \text{Ro}(\mathcal{R}') = \emptyset$: Da $T_{\mathcal{R}}(Y'')$ in \mathcal{R}' ersetzt worden ist, folgt $\text{Ro}_{\mathcal{R}'}(Y'') \cap \text{Ro}_{\mathcal{R}'}(\mathcal{R}') = \emptyset$. Da nach Voraussetzung $Z'' \subseteq Y''$ gilt, folgt $\text{Ro}(Z'') \cap \text{Ro}(\mathcal{R}') = \emptyset$.
4. $B(Z'') \times p \in \mathcal{R}'$: Nach Voraussetzung gilt $Z'' \subseteq Y''$ und damit folgt $B(Z'') \times p \in T_{\mathcal{S}}((B(Y'')) \times p)$. Da $T_{\mathcal{S}}((B(Y'')) \times p)$ der ersetzte Teilbaum in \mathcal{R}' ist, folgt die Behauptung.
5. $B(\text{Ro}(T_{\mathcal{R}}(Z''))) = B(B(Z'') \times p)$: $B(\text{Ro}(T_{\mathcal{R}}(Z''))) = B(Z'') = B(B(Z'') \times p)$.

□

Da \mathcal{R}'' eine Ersetzung ist, folgt aus Theorem 4.1, daß \mathcal{R}'' ein m -Clustersystem ist. Damit ist die erste Bedingung für $\mathcal{R}'' \in \text{RC}(\mathcal{G}, \mathcal{S})$ gezeigt.

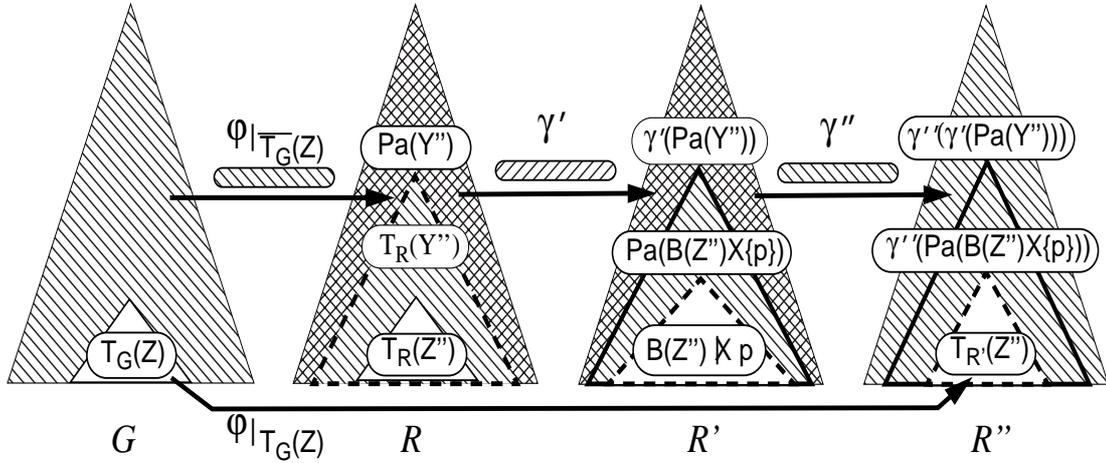
Aussage 2 Für alle $V, W \in \mathcal{R}''$ mit $V \subseteq W$ gilt $B(K_{\mathcal{R}''}(W, V)) = K_{\mathcal{S}}(B(W), B(V))$.

Beweis. Wir zeigen zuerst $B(K_{\mathcal{R}'}(W, V)) = K_{\mathcal{S}}(B(W), B(V))$, für alle $V, W \in \mathcal{R}'$ mit $V \subseteq W$. Aus Lemma 9.1 folgt, daß \mathcal{R}' eine Ersetzung ist, und es gilt $B(K_{B(Y'') \times p}(W, V)) = K_{\mathcal{S}}(B(W), B(V))$ für alle $V, W \in B(Y'') \times p$ mit $V \subseteq W$. Damit ist Lemma 4.5 anwendbar, woraus die Behauptung folgt.

Wir zeigen nun die eigentliche Aussage, indem wir wieder die Anwendbarkeit von Lemma 4.5 zeigen. Nach Definition des Reconciled Tree gilt $B(K_{\mathcal{R}}(W, V)) = K_{\mathcal{S}}(B(W), B(V))$ für alle $V, W \in \mathcal{R}$ mit $V \subseteq W$, und damit gilt diese Aussage auch für jeden Teilbaum von \mathcal{R} , also auch für $T_{\mathcal{R}}(Z'')$. Mit der vorherigen Aussage für \mathcal{R}' ist nun Lemma 4.5 anwendbar, woraus nun die Behauptung folgt. □

Wir werden nun zeigen, daß ein $\varphi'' \in \text{Emon}(\mathcal{G}, \mathcal{R}'')$ existiert. φ'' werden wir aus den Funktionen φ sowie den durch die Ersetzungen \mathcal{R}' und \mathcal{R}'' definierten Funktionen $\Upsilon' := \Upsilon_{\mathcal{R}, T_{\mathcal{S}}(B(Y'')) \times p, Y''}$ und $\Upsilon'' := \Upsilon_{\mathcal{R}', T_{\mathcal{R}}(\varphi(Z)), B(Z'') \times \{p\}}$ konstruieren (siehe Abbildung 9.2).

Dazu partitionieren wir die Urbildmenge von φ in $T_{\mathcal{G}}(Z)$ und $\overline{T_{\mathcal{G}}}(Z)$. Die partielle Abbildung $\varphi|_{T_{\mathcal{R}}(Z)}$ übernehmen wir direkt für die partielle Abbildung $\varphi''|_{T_{\mathcal{R}}(Z)}$. Dagegen wird $\overline{T_{\mathcal{G}}}(Z)$ durch $\Upsilon'' \circ \Upsilon' \circ \varphi|_{\overline{T_{\mathcal{G}}}(Z)}$ von \mathcal{R} über \mathcal{R}' nach \mathcal{R}'' abgebildet. Um zu zeigen, daß $\varphi'' \in \text{Emon}(\mathcal{G}, \mathcal{R}'')$ ist, zeigen wir folgendes:

Abbildung 9.2: Konstruktion der Abbildung φ''

Aussage 3 $\gamma'' \circ \gamma' \circ \varphi|_{\overline{T_G(Z)}} \in \text{Emon}(\overline{T_G(Z)}, \overline{T_{R''}(Z'')})$.

Beweis. Wir zeigen $\gamma'' \circ \gamma' \in \text{Emon}(\overline{T_R(Y'')}, \overline{T_{R''}(Z'')})$. Nach Korollar 4.1 gilt, $\gamma' \in \text{Eiso}(\overline{T_R(Y'')}, \overline{T_{R'}(B(Y'') \times \{p\})})$ und $\gamma'' \in \text{Eiso}(\overline{T_{R'}(B(Z'') \times \{p\})}, \overline{T_{R''}(Z'')})$. Nach Voraussetzung gilt $Z'' \subseteq Y''$, woraus $B(Z'') \times \{p\} \subseteq B(Y'') \times \{p\}$ folgt. Also ist $\overline{T_{R'}(B(Y'') \times \{p\})} \subseteq \overline{T_{R'}(B(Z'') \times \{p\})}$, und damit $\gamma'' \circ \gamma' \in \text{Emon}(\overline{T_R(Y'')}, \overline{T_{R''}(Z'')})$.

Wir zeigen die eigentliche Aussage. Nach Voraussetzung gilt $\varphi \in \text{Emon}(\mathcal{G}, \mathcal{R})$. Da $\overline{T_G(Z)}$ gegen \sqcup abgeschlossen ist, folgt $\varphi|_{\overline{T_G(Z)}} \in \text{Emon}(\overline{T_G(Z)}, \overline{T_{\mathcal{R}}(Z'')})$. Nach Voraussetzung ist Z das einzige Kind von X mit $\varphi(Z) \subseteq Y''$, woraus $\varphi|_{\overline{T_G(Z)}} \in \text{Emon}(\overline{T_G(Z)}, \overline{T_{\mathcal{R}}(Y'')})$ folgt. Mit $\gamma'' \circ \gamma' \in \text{Emon}(\overline{T_{R'}(Y'')}, \overline{T_{R''}(\varphi(Z))})$ folgt die Aussage. \square

Mit $\gamma'' \circ \gamma' \circ \varphi|_{\overline{T_G(Z)}}$ können wir nun die Elemente aus $\overline{T_G(Z)}$ in $\overline{T_{R''}(\varphi(Z))}$ einbetten. Um die verbleibenden Elemente $T_G(Z)$ in \mathcal{R}'' einzubetten, verwenden wir die partielle Funktion $\varphi|_{T_G(Z)}: T_G(Z) \rightarrow T_{\mathcal{R}}(Z)$. Denn erstens ist $T_G(Z)$ gegenüber \sqcup abgeschlossen und damit $\varphi|_{T_G(Z)} \in \text{Emon}(T_G(Z), T_{\mathcal{R}}(Z''))$, und zweitens ist $T_{\mathcal{R}}(Z'') = T_{\mathcal{R}''}(Z'')$, womit $\varphi|_{T_G(Z)} \in \text{Emon}(T_G(Z), T_{\mathcal{R}''}(Z''))$ gilt. Die beiden partiellen Funktionen $\gamma'' \circ \gamma' \circ \varphi|_{\overline{T_G(Z)}} \in \text{Emon}(\overline{T_G(Z)}, \overline{T_{R''}(Z'')})$ und $\varphi|_{T_G(Z)}$ bilden in disjunkte Mengen aus \mathcal{R}'' ab, womit wir die folgende Funktion definieren können.

Festlegung 5 Sei $\varphi'': \mathcal{G} \rightarrow \mathcal{R}''$ mit

$$\varphi''(P) := \begin{cases} \gamma'' \circ \gamma' \circ \varphi|_{\overline{T_G(Z)}} & P \in \overline{T_G(Z)} \\ \varphi|_{T_G(Z)} & \text{sonst.} \end{cases}$$

Es bleibt $\varphi'' \in \text{Emon}(\mathcal{G}, \mathcal{R}'')$ zu zeigen. Da φ'' aus den Einbettungsmonomorphismen $\Upsilon'' \circ \Upsilon' \circ \varphi|_{\overline{T_{\mathcal{G}}(Z)}}$ und $\varphi|_{T_{\mathcal{G}}(Z)}$ zusammengesetzt ist, bleibt die Einbettungsmonomorphie nur noch für $A \in T_{\mathcal{G}}(Z)$ und $B \in \overline{T_{\mathcal{G}}(Z)}$ zu zeigen.

Aussage 4 Seien $A \in T_{\mathcal{G}}(Z)$ und $B \in \overline{T_{\mathcal{G}}(Z)}$, dann gilt $\varphi''(A \sqcup B) = \varphi''(A) \sqcup \varphi''(B)$.

Beweis. Wir zeigen zuerst: $\varphi''(A \sqcup B) = \varphi''(X) \sqcup \varphi''(B)$.

$$\begin{aligned} A \sqcup B &= \min(K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), B) \cap K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), A)) \\ &= \min(K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), B) \cap (K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), X) \dot{\cup} K_{\mathcal{G}}(Z, A))) \\ &= \min(K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), B) \cap K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), X)) \\ &= A \sqcup X \end{aligned} \tag{9.1}$$

Zu (9.1): Aus der Voraussetzung $A \in T_{\mathcal{G}}(Z)$ folgt $K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), B) \subseteq \overline{T_{\mathcal{G}}(Z)}$, und aus der Voraussetzung $B \in \overline{T_{\mathcal{G}}(Z)}$ folgt $K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), A) = K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), X) \dot{\cup} K_{\mathcal{G}}(Z, A)$ mit $K_{\mathcal{G}}(Z, A) \subseteq T_{\mathcal{G}}(Z)$ sowie $K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), X) \subset \overline{T_{\mathcal{G}}(Z)}$. Hieraus folgt dann $K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), B) \cap K_{\mathcal{G}}(Z, A) = \emptyset$.

Nach Aussage 3 gilt $\varphi''|_{\overline{T_{\mathcal{G}}(Z)}} \in \text{Emon}(\overline{T_{\mathcal{G}}(Z)}, \overline{T_{\mathcal{R}''}(Z'')})$, und nach Voraussetzung gilt $A, X \in \overline{T_{\mathcal{G}}(Z)}$. Damit folgt $\varphi''(A \sqcup B) = \varphi''(X \sqcup B) = \varphi''(X) \sqcup \varphi''(B)$.

Wir zeigen: $\varphi''(X) \sqcup \varphi''(B) = \varphi''(A) \sqcup \varphi''(B)$. Wir zeigen hierzu $Z'' \subset \varphi''(X)$. Da nach Voraussetzung $Y'' \subset X''$ gilt, folgt $\Upsilon'(X'') = X'' \setminus B(Y'') \cup B(Y'') \times \{p\}$. Wegen $B(Z'') \times \{p\} \subset \Upsilon'(X'')$ folgt $\Upsilon''(\Upsilon'(X'')) = \Upsilon'(X'') \setminus B(Z'') \times \{p\} \cup Z''$. Somit gilt die Aussage.

Sei $C \in \text{Ch}_{\mathcal{R}''}(\varphi''(X))$ mit $Z'' \subseteq C$, dann gilt:

$$\begin{aligned} \varphi''(A) \sqcup \varphi''(B) &= \min(K_{\mathcal{R}''}(\text{Ro}(\mathcal{R}''), \varphi''(A)) \cap K_{\mathcal{R}''}(\text{Ro}(\mathcal{R}''), \varphi''(B))) \\ &= \min((K_{\mathcal{R}''}(\text{Ro}(\mathcal{R}''), \varphi''(X)) \dot{\cup} K_{\mathcal{R}''}(C, \varphi''(A))) \\ &\quad \cap K_{\mathcal{R}''}(\text{Ro}(\mathcal{R}''), \varphi''(B))) \end{aligned} \tag{9.2}$$

$$\begin{aligned} &= \min(K_{\mathcal{R}''}(\text{Ro}(\mathcal{R}''), \varphi''(X)) \cap K_{\mathcal{R}''}(\text{Ro}(\mathcal{R}''), \varphi''(B))) \\ &= \varphi''(X) \sqcup \varphi''(B) \end{aligned} \tag{9.3}$$

Zu (9.2): Da nach Voraussetzung $A, Z \in T_{\mathcal{G}}(Z)$ gilt, folgt $\varphi''(A) = \varphi(A)$ und $\varphi''(Z) = \varphi(Z)(= Z'')$. Aus der Voraussetzung $A \subseteq Z$ folgt damit $\varphi''(A) \subseteq Z''$. Wir haben $Z'' \subseteq C$ gezeigt, womit die Kette $K_{\mathcal{R}''}(\text{Ro}(\mathcal{R}''), \varphi''(A))$ über C läuft.

Zu (9.3): Wir zeigen hierzu $B \notin K_{\mathcal{R}''}(C, \varphi''(A))$.

Nehmen wir $\varphi''(B) \in K_{\mathcal{R}''}(C, \varphi''(A))$ an. Da nach Voraussetzung $B \in T_{\mathcal{G}}(Z)$ ist, folgt $\varphi''(B) \in K_{\mathcal{R}''}(C, \text{Pa}_{\mathcal{R}''}(Z''))$. Wegen $Z'' \subset \varphi''(B)$ gilt $B(Z'') \times \{p\} \subset$

$(\Upsilon'')^{-1}(\varphi''(B))$. Aus $(\Upsilon') : \overline{T_{\mathcal{R}}(Y'')} \rightarrow \overline{T_{\mathcal{R}'}(B(Y'') \times \{p\})}$ folgt $B(Y'') \times \{p\} \subset (\Upsilon'')^{-1}(\varphi''(B))$ und damit $Y'' \subset (\Upsilon')^{-1}((\Upsilon'')^{-1}(\varphi''(B))) = \varphi(B)$.

Da nach Aussage 3 $\Upsilon'' \circ \Upsilon' \in \text{Emon}(\overline{T_{\mathcal{R}'}(B(Y''))}, \overline{T_{\mathcal{R}''}(Z'')})$ gilt, folgt $X \subset B$ aus der Annahme $\varphi''(B) \subset \varphi''(X)$.

Somit gilt $\varphi(Z) \subseteq Y'' \subset \varphi(B) \subset \varphi(X)$, und mit der Voraussetzung $\varphi \in \text{Emon}(\mathcal{G}, \mathcal{R})$ folgt $Z \subset B \subset X$. Dies steht im Widerspruch zu der Voraussetzung $Z \in \text{Ch}_{\mathcal{G}}(X)$.

Abschließend gilt damit $\varphi''(A \sqcup B) = \varphi''(X) \sqcup \varphi''(B) = \varphi''(A) \sqcup \varphi''(B)$. \square

Es verbleibt nun noch die minimale Kardinalität von \mathcal{R}'' zu zeigen. Da nach Voraussetzung \mathcal{R} minimal ist, zeigen wir dies durch die folgende Aussage.

Aussage 5 $|\mathcal{R}''| = |\mathcal{R}|$.

Beweis. Es gilt

$$\begin{aligned} |\mathcal{R}''| &= (|\mathcal{R}| - |T_{\mathcal{R}}(Y'')| + |T_{\mathcal{S}}(B(Y'')) \times p| - |T_{\mathcal{S}}(B(Z'')) \times p| + |T_{\mathcal{R}}(Z'')|) \\ &= |\mathcal{R}| - |T_{\mathcal{R}}(Y'')| + |T_{\mathcal{S}}(B(Y''))| - |T_{\mathcal{S}}(B(Z''))| + |T_{\mathcal{R}}(Z'')| \\ &= |\mathcal{R}| - (|T_{\mathcal{R}}(Y'')| - |T_{\mathcal{R}}(Z'')| - |T_{\mathcal{S}}(B(Y''))| + |T_{\mathcal{S}}(B(Z''))|) \\ &= |\mathcal{R}| - c \end{aligned}$$

mit $c := |T_{\mathcal{R}}(Y'')| - |T_{\mathcal{R}}(Z'')| - |T_{\mathcal{S}}(B(Y''))| + |T_{\mathcal{S}}(B(Z''))|$.

Es gilt $|T_{\mathcal{R}}(Y'')| - |T_{\mathcal{R}}(Z'')| \geq |T_{\mathcal{S}}(B(Y''))| - |T_{\mathcal{S}}(B(Z''))|$, woraus $c \geq 0$ folgt. Somit ist $|\mathcal{R}''| = |\mathcal{R}| - c$ mit $c \in \mathbb{N}_0$. Da \mathcal{R} minimal ist, folgt $c = 0$, also $|\mathcal{R}''| = |\mathcal{R}|$. \square

\square

Verw. von: Theo. 4.1, Kor. 4.1, Lem. 9.1, 4.5; *Verw. in:* Lem. 4.8.

Lemma 9.3

Sei $X \in \mathcal{G}$ und $\{Z_1, \dots, Z_n\} := \text{Ch}_{\mathcal{G}}(X)$ die Kinder von X , sodaf

$$B(\varphi(Z_i)) \sqcup B(\varphi(Z_j)) = B(\varphi(X)) \text{ und } B(\varphi(Z_i)) \subset B(\varphi(X)),$$

für alle $i, j \in \{1, \dots, n\}$ mit $i \neq j$ gilt.

Für ein $p \in (\mathbb{N})$, welches tdisjunkt zu $\text{Ro}(\mathcal{R})$ ist, sei

$$\mathcal{R}^0 := \text{Repl}(\mathcal{R}, T_{\mathcal{S}}(B(\varphi(X))) \times p, \varphi(X)) \text{ und}$$

$$\mathcal{R}^i := \text{Repl}(\mathcal{R}^{i-1}, T_{\mathcal{R}}(\varphi(Z_i)), B(\varphi(Z_i)) \times \{p\}) \text{ für alle } i \in \{1, \dots, n\}.$$

Dann ist $\mathcal{R}^n \in \text{RC}(\mathcal{G}, \mathcal{S})$.

Beweis. Zur Vereinfachung bezeichnen wir die Ersetzungen mit definierenden Funktionen wie folgt.

Festlegung 1 *Wir definieren für alle $i \in \{1, \dots, n\}$:*

$$\Upsilon_0 := \Upsilon_{\mathcal{R}, B(T_{\mathcal{R}}(\varphi(X))) \times p, \varphi(X)}, \quad \Upsilon_i := \Upsilon_{\mathcal{R}^{i-1}, T_{\mathcal{R}}(\varphi(Z_i)), B(\varphi(Z_i)) \times \{p\}}$$

Weiter bezeichnen wir $X'' := \varphi(X)$ und $Z_i'' := \varphi(Z_i)$ für alle $i \in \{1, \dots, n\}$.

Zum Beweis zeigen wir zuerst, \mathcal{R}^i ist die Ersetzung von $T_{\mathcal{R}}(Z_i'')$ an der Stelle $B(Z_i) \times \{p\}$ in \mathcal{R}^{i-1} für alle $i \in \{1, \dots, n\}$. Diese Aussage finden wir in Aussage 1 (1) wieder. Punkte (2) und (3) dienen dazu, die nächste Ersetzung zu garantieren.

Aussage 1 *Für alle $i \in \{1, \dots, n\}$ gilt:*

(1): \mathcal{R}^i ist die Ersetzung von $T_{\mathcal{R}}(Z_i'')$ an Element $B(Z_i'') \times \{p\}$ in \mathcal{R}^{i-1}

(2): $\text{Ro}(\mathcal{R}^i) \cap \bigcup_{j \in \{i+1, \dots, n\}} Z_j'' = \emptyset$

(3): $\{Z_1'', \dots, Z_i'', B(Z_{i+1}'') \times \{p\}, \dots, B(Z_n'') \times \{p\}\} \subseteq \mathcal{R}^i$

Beweis. Wir zeigen dies durch v.I. nach $i \in \{1, \dots, n\}$.

$i = 0$: Nach Lemma 9.1 ist \mathcal{R}^0 die Ersetzung von $T_{\mathcal{S}}(B(X'')) \times p$ an Element X'' in \mathcal{R} . Also ist $T_{\mathcal{S}}(B(X'')) \times p \subseteq \mathcal{R}^0$ und damit $\{B(Z_1'') \times \{p\}, \dots, B(Z_n'') \times \{p\}\} \subseteq \mathcal{R}^0$. Da an Element X'' ersetzt wurde, ist $T_{\mathcal{R}}(X'') \notin \mathcal{R}^0$ und damit $\text{Ro}(\mathcal{R}^0) \cap \bigcup_{j \in \{1, \dots, n\}} Z_j'' = \emptyset$.

$i \rightarrow i + 1$: I.V: Es gelte die Behauptung für $i \in \{0, \dots, n - 1\}$.

Wir zeigen die Behauptung:

(1): Nach I.V. (1) ist \mathcal{R}^i eine Ersetzung und mit Theorem 4.1 folgt, daß \mathcal{R}^i ein m -Clustersystem ist. Aus der I.V. (3) folgt $B(Z_{i+1}'') \times \{p\} \in \mathcal{R}^i$, also kann an diesem Element in \mathcal{R}^i ersetzt werden. Dies kann durch $T_{\mathcal{R}}(Z_{i+1}'')$ erfolgen, da nach I.V. (2) $\text{Ro}(\mathcal{R}^i) \cap Z_{i+1}'' = \emptyset$ gilt, und $B(\text{Ro}(T_{\mathcal{R}}(\varphi(Z_{i+1})))) = B(Z_{i+1}'') = B(B(Z_{i+1}'') \times \{p\})$ ist. Somit ist \mathcal{R}^{i+1} eine Ersetzung.

(2):

$$\begin{aligned}
& \text{Ro}(\mathcal{R}^{i+1}) \cap \bigcup_{j \in \{i+2, \dots, n\}} Z_j'' \\
&= (\text{Ro}(\mathcal{R}^i) \setminus B(Z_{i+1}'') \times \{p\}) \cup T_{\mathcal{R}}(Z_{i+1}'') \cap \bigcup_{j \in \{i+2, \dots, n\}} Z_j'' \quad (9.4) \\
&\subseteq (\text{Ro}(\mathcal{R}^i) \cup T_{\mathcal{R}}(Z_{i+1}'')) \cap \bigcup_{j \in \{i+2, \dots, n\}} Z_j'' \\
&= \text{Ro}(\mathcal{R}^i) \cap \bigcup_{j \in \{i+2, \dots, n\}} Z_j'' \\
&= \emptyset \quad (9.5)
\end{aligned}$$

Zu (9.4): Gilt, da, wie eben gezeigt, \mathcal{R}^{i+1} eine Ersetzung ist.

Zu (9.5): Gilt nach I.V. (2).

(3): Da \mathcal{R}^{i+1} eine Ersetzung ist, gilt

$$\begin{aligned}
\mathcal{R}^{i+1} &= \Upsilon^{i+1}(\mathcal{R}^i \setminus T_{\mathcal{R}^i}(B(Z_{i+1}'') \times \{p\})) \cup T_{\mathcal{R}}(Z_{i+1}'') \\
&= (\mathcal{R}^i \setminus B(\varphi(Z_{i+1}'')) \times \{p\} \uplus \varphi(Z_{i+1}'')) \cup T_{\mathcal{R}}(Z_{i+1}'').
\end{aligned}$$

\mathcal{R}^{i+1} besteht somit aus dem Teilbaum $T_{\mathcal{R}}(Z_{i+1}'')$ und den Mengen aus \mathcal{R}^i , aus denen jeweils die Elemente aus $B(Z_{i+1}'') \times \{p\}$ entfernt sowie die Elemente aus Z_{i+1}'' hinzugefügt wurden.

Nach I.V. (3) gilt $\{Z_1'', \dots, Z_i'', B(Z_{i+1}'') \times \{p\}, \dots, B(Z_n'') \times \{p\}\} \subseteq \mathcal{R}^i$. Wir stellen nun sicher, daß mit $B(Z_{i+1}'') \times \{p\}$ keine Elemente aus $B(Z_m'') \times \{p\}$ oder aus Z_m'' , mit $m \in \{i+2, \dots, n\}$ und $m' \in \{1, \dots, i\}$ aus Mengen von \mathcal{R}^i entfernt werden. Aus der Voraussetzung

$$B(Z_k'') \cup B(Z_l'') = B(X'') \text{ und } B(Z_l'') \subset B(X''),$$

für alle $k, l \in \{1, \dots, n\}$ mit $k \neq l$, folgt $B(Z_m'') \times \{p\} \not\subseteq B(Z_{i+1}'') \times \{p\}$, für alle $m \in \{i+2, \dots, n\}$. Da p tdisjunkt zu $\text{Ro}(\mathcal{R})$ ist, folgt $Z_m'' \not\subseteq B(Z_{i+1}'') \times \{p\}$, für alle $m' \in \{1, \dots, i+1\}$.

Somit folgt $\{Z_1'', \dots, Z_{i+1}'', B(Z_{i+2}'') \times \{p\}, \dots, B(Z_n'') \times \{p\}\} \subseteq \mathcal{R}^{i+1}$.

□

Wir werden, ähnlich wie in Lemma 9.2, zeigen, daß eine Abbildung $\varphi_{\text{neu}} \in \text{Emon}(\mathcal{G}, \mathcal{R})$ existiert. φ_{neu} werden wir durch die Funktionen $\Upsilon_0, \dots, \Upsilon_n$, welche durch die Ersetzungen $\mathcal{R}^0, \dots, \mathcal{R}^n$ definiert sind, konstruieren (siehe Abbildung 4.10). Dazu partitionieren wir die Urbildmenge von φ in $\mathcal{Z} := T_{\mathcal{G}}(X) \setminus \{X\}$ und $\bar{\mathcal{Z}} := \mathcal{G} \setminus \mathcal{Z}$. Da $\varphi(\mathcal{Z}) \subset \mathcal{R}^n$ nach Aussage 1 gilt, übernehmen wir die partielle

Abbildung $\varphi|_{\mathcal{Z}}$ für $\varphi_{\text{neu}}|_{\mathcal{Z}}$. Hingegen $\varphi_{\text{neu}}|_{\overline{\mathcal{Z}}}$ bilden wir über $\Upsilon_n \circ \dots \circ \Upsilon_0 \circ \varphi|_{\overline{\mathcal{Z}}}$ nach \mathcal{R}^n ab.

Dazu zeigen wir zuerst folgende Aussage:

Aussage 2 Seien $\Upsilon^0 := \Upsilon_0 \cup \{(\varphi(X), B(\varphi(X)) \times \{p\})\}$ und $\Upsilon^i := \Upsilon_i \circ \Upsilon^{i-1}$ für jedes $i \in \{1, \dots, n\}$. Dann gilt

1. $\Upsilon^i : \overline{\mathcal{T}_{\mathcal{R}}}(X'') \rightarrow \mathcal{R}^i$ ist eine Funktion,
mit $\Upsilon^i \in \text{Emon}(\overline{\mathcal{T}_{\mathcal{R}}}(X'') \cup \{X''\}, \overline{\mathcal{T}_{\mathcal{R}^i}}(\Upsilon^i(X'')) \cup \{\Upsilon^i(X'')\})$.
2. $M \subset \Upsilon^i(X)$ und $M \sqcup_{\mathcal{R}^i} N = \Upsilon^i(X)$, für alle
 $M, N \in \{Z_1'', \dots, Z_i'', B(Z_{i+1}'') \times \{p\}, \dots, B(Z_n'') \times \{p\}\}$ mit $M \neq N$.

für jedes $i \in \{0, \dots, n\}$.

Beweis. Wir zeigen dies durch V.I. nach $i \in \{0, \dots, n\}$.

$i = 0$: Wir zeigen die Behauptung:

- (i): Wir zeigen zuerst, daß Υ^0 eine Funktion ist. Nach Voraussetzung ist $\Upsilon^0 = \Upsilon_0 \cup \{(X'', B(X'') \times \{p\})\}$. Nach Aussage 1 ist \mathcal{R}^0 eine Ersetzung und damit $\Upsilon_0 : \overline{\mathcal{T}_{\mathcal{R}}}(X'') \rightarrow \overline{\mathcal{T}_{\mathcal{R}^0}}(B(X'') \times \{p\})$ als Funktion definiert. Somit ist $\Upsilon^0 : \overline{\mathcal{T}_{\mathcal{R}}}(X'') \cup \{X''\} \rightarrow \overline{\mathcal{T}_{\mathcal{R}^0}}(B(X'') \times \{p\}) \cup \{B(X'') \times \{p\}\}$ eine Funktion.

Wir zeigen nun $\Upsilon^0 \in \text{Emon}(\overline{\mathcal{T}_{\mathcal{R}}}(X'') \cup \{X''\}, B(X'') \times \{p\} \cup \{(B(X'') \times \{p\})\})$. Aus Lemma 4.2 folgt, $\Upsilon_0 \in \text{Oiso}(\overline{\mathcal{T}_{\mathcal{R}}}(X''), \overline{\mathcal{T}_{\mathcal{R}^0}}(B(X'') \times \{p\}))$. Hieraus folgt die Bijektivität von Υ^0 . Es bleibt die Ordnungshomomorphie von Υ^0 für das Element X'' zu zeigen. Sei $Y'' \in \overline{\mathcal{T}_{\mathcal{R}}}(X'')$. Wir zeigen die Ordnungshomomorphie für die beiden folgenden Fälle:

1. $Y'' \cap X'' = \emptyset$: Es gilt also $\Upsilon^0(Y'') = Y''$ und $\Upsilon^0(X'') = B(X'') \times \{p\}$.
Da p tdisjunkt zu $\text{Ro}(\mathcal{R})$ ist, folgt $B(X'') \times \{p\} \cap Y'' = \emptyset$ und somit $\Upsilon^0(X'') \cap \Upsilon^0(Y'') = \emptyset$.
2. $X'' \subset Y''$: Es gilt
 $\Upsilon(X'') = B(X'') \times \{p\} \subset Y'' \setminus X'' \cup B(X'') \times \{p\} = \Upsilon(Y'')$.

Somit folgt

$$\begin{aligned} \Upsilon^0 &\in \text{Oiso}(\overline{\mathcal{T}_{\mathcal{R}}}(X'') \cup \{X''\}, \overline{\mathcal{T}_{\mathcal{R}^0}}(B(X'') \times \{p\}) \cup \{(B(X'') \times \{p\})\}) \\ &= \text{Oiso}(\overline{\mathcal{T}_{\mathcal{R}}}(X'') \cup \{X''\}, \overline{\mathcal{T}_{\mathcal{R}^0}}(\Upsilon^0(X'')) \cup \{\Upsilon^0(X'')\}) \quad (9.6) \\ &= \text{Eiso}(\overline{\mathcal{T}_{\mathcal{R}}}(X'') \cup \{X''\}, \overline{\mathcal{T}_{\mathcal{R}^0}}(\Upsilon^0(X'')) \cup \{\Upsilon^0(X'')\}) \end{aligned}$$

Zu (9.6): Nach Definition von Υ^0 gilt $\Upsilon^0(X'') = B(X'') \times \{p\}$.

(ii): Nach Voraussetzung gilt $B(Z''_k) \sqcup B(Z''_l) = B(X'')$ und $B(Z''_k) \subset B(X'')$, für alle $k, l \in \{1, \dots, n\}$ mit $k \neq l$. Hieraus folgt $B(Z''_k) \times \{p\} \sqcup B(Z''_l) \times \{p\} = B(X'') \times \{p\} = \Upsilon^0(X'')$ und $B(Z''_k) \subset B(X'') \times \{p\} = \Upsilon^0(X'')$, für alle $k, l \in \{1, \dots, n\}$ mit $k \neq l$.

$i \rightarrow i + 1$: I.V.: Es gelte die Behauptung für $i \in \{1, \dots, n - 1\}$.

Wir zeigen die Behauptung:

(i): Wir zeigen zuerst, daß Υ^{i+1} eine Funktion ist. Es gilt nach Definition $\Upsilon^{i+1} = \Upsilon_{i+1} \circ \Upsilon^i$. Für Υ_{i+1} und Υ^i gilt:

$\Upsilon_{i+1} : \overline{\mathbb{T}_{\mathcal{R}^i}}(B(Z''_{i+1}) \times \{p\}) \rightarrow \overline{\mathbb{T}_{\mathcal{R}^{i+1}}}(Z''_{i+1})$ ist eine Funktion, da nach Aussage 1 \mathcal{R}^{i+1} eine Ersetzung ist.

I.V. (1) besagt, $\Upsilon : \overline{\mathbb{T}_{\mathcal{R}}}(X'') \cup \{X''\} \rightarrow \mathcal{R}^i$ eine Funktion, mit $\Upsilon^i \in \text{Emon}(\overline{\mathbb{T}_{\mathcal{R}}}(X'') \cup \{X''\}, \overline{\mathbb{T}_{\mathcal{R}^i}}(\Upsilon^i(X'')) \cup \{\Upsilon^i(X'')\})$,

$\Upsilon_{i+1} \circ \Upsilon^i$ ist eine Funktion, falls die Bildmenge von Υ^i in der Urbildmenge von Υ_{i+1} enthalten ist. Aus I.V. (2) folgt $(B(Z''_{i+1}) \times \{p\}) \subset \Upsilon^i(X'')$, und damit gilt $\overline{\mathbb{T}_{\mathcal{R}^i}}(\Upsilon^i(X'')) \cup \{\Upsilon^i(X'')\} \subset \overline{\mathbb{T}_{\mathcal{R}^i}}(B(Z''_{i+1}) \times \{p\})$. Somit ist $\Upsilon^{i+1} : B(Z''_{i+1}) \times \{p\} \rightarrow \overline{\mathbb{T}_{\mathcal{R}^{i+1}}}(Z''_{i+1})$ eine Funktion.

Wir zeigen die Einbettungsmorphismie von Υ^{i+1} . Es gilt:

$\Upsilon^i \in \text{Emon}(\overline{\mathbb{T}_{\mathcal{R}}}(X'') \cup \{X''\}, \overline{\mathbb{T}_{\mathcal{R}^i}}(X'') \cup \{\Upsilon^i(X'')\})$ nach I.V. (1).

$\Upsilon_{i+1} \in \text{Eiso}(B(Z''_{i+1}) \times \{p\}, \overline{\mathbb{T}_{\mathcal{R}^{i+1}}}(Z''_{i+1}))$ folgt aus Korollar 4.1.

Damit ist $\Upsilon^{i+1} \in \text{Emon}(\overline{\mathbb{T}_{\mathcal{R}}}(X'') \cup \{X''\}, \overline{\mathbb{T}_{\mathcal{R}^{i+1}}}(Z''_{i+1}))$.

Wir formulieren nun noch die Bildmenge von Υ^{i+1} genauer. Da

$X'' = \min(\overline{\mathbb{T}_{\mathcal{R}}}(X'') \cup \{X''\})$ und

$\Upsilon^{i+1} \in \text{Emon}(\overline{\mathbb{T}_{\mathcal{R}}}(X'') \cup \{X''\}, \overline{\mathbb{T}_{\mathcal{R}^{i+1}}}(Z''_{i+1}))$,

folgt $\Upsilon^{i+1}(X'') = \min(\Upsilon^{i+1}(\overline{\mathbb{T}_{\mathcal{R}}}(X'') \cup \{X''\}))$. Somit folgt

$\Upsilon^{i+1} \in \text{Emon}(\overline{\mathbb{T}_{\mathcal{R}}}(X'') \cup \{X''\}, \overline{\mathbb{T}_{\mathcal{R}^{i+1}}}(\Upsilon^{i+1}(X'')) \cup \{\Upsilon^{i+1}(X'')\})$.

(ii): Wir zeigen die Behauptung zuerst für

$M, N \in \{Z''_1, \dots, Z''_i, B(Z''_{i+2}) \times \{p\}, \dots, B(Z''_n) \times \{p\}\}$.

Wir zeigen $\Upsilon_{i+1}(M) = M$ und $\Upsilon_{i+1}(N) = N$. Aus Aussage 1 folgt $M, N \in \mathcal{R}^i$. Nach I.V. (2) gilt $M \subset \Upsilon^i(X'')$, $B(Z''_{i+1}) \times \{p\} \subset \Upsilon^i(X'')$ und $M \sqcup B(Z''_{i+1}) \times \{p\} = \Upsilon^i(X'')$. Hieraus folgt $M \cap B(Z''_{i+1}) \times \{p\} = \emptyset$. Entsprechend folgt $N \cap B(Z''_{i+1}) \times \{p\} = \emptyset$. Somit sind M, N Elemente der Urbildmenge $\overline{\mathbb{T}_{\mathcal{R}^i}}(B(Z''_{i+1}) \times \{p\})$ von Υ_{i+1} mit $\Upsilon_{i+1}(M) = M$ und $\Upsilon_{i+1}(N) = N$.

Weiter gilt nach Korollar 4.1, daß $\Upsilon_{i+1} \in \text{Eiso}(\mathbf{B}(Z''_{i+1}) \times \{p\}, \overline{\mathbf{T}_{\mathcal{R}^{i+1}}}(Z''_{i+1}))$ und somit auch ordnungsisomorph ist.

Damit folgt aus der I.V. (2), $M \subset \Upsilon^i(X'')$ die Aussage

$$M = \Upsilon_{i+1}(M) \subset \Upsilon_{i+1}(\Upsilon^i(X'')) = \Upsilon^{i+1}(X'').$$

Aus der I.V. (2), $M \sqcup N = \Upsilon^i(X'')$ folgt

$$M \sqcup N = \Upsilon_{i+1}(M) \sqcup \Upsilon_{i+1}(N) = \Upsilon_{i+1}(M \sqcup N) = \Upsilon_{i+1}(\Upsilon^i(X'')) = \Upsilon^{i+1}(X'')$$

Wir zeigen nun die verbleibenden Fälle

$$Z''_{i+1} \subset \Upsilon^{i+1}(X'') \text{ und } M \sqcup Z''_{i+1} = \Upsilon^{i+1}(X''),$$

für alle $M \in \{Z''_1, \dots, Z''_i, \mathbf{B}(Z''_{i+2}) \times \{p\}, \dots, \mathbf{B}(Z''_n) \times \{p\}\}$.

Wir zeigen $Z''_{i+1} \subset \Upsilon^{i+1}(X'')$. Nach I.V. gilt $\mathbf{B}(Z''_{i+1} \times \{p\}) \subset \Upsilon^i(X'')$, und damit $\Upsilon_{i+1}(\Upsilon^i(X'')) = \Upsilon^i(X'') \setminus \mathbf{B}(Z''_{i+1}) \times \{p\} \cup Z''_{i+1}$. Somit folgt $Z''_{i+1} \subset \Upsilon_{i+1}(\Upsilon^i(X'')) = \Upsilon^{i+1}(X'')$.

Wir zeigen $M \sqcup Z''_{i+1} = \Upsilon^{i+1}(X'')$. Nach der I.V. (2) gilt

$$M \subset \Upsilon^i(X''), \mathbf{B}(Z''_{i+1}) \times \{p\} \subset \Upsilon^i(X'') \text{ und}$$

$$M \sqcup \mathbf{B}(Z''_{i+1}) \times \{p\} = \Upsilon^i(X'').$$

Es folgt hieraus $M \sqcup \text{Pa}_{\mathcal{R}^i}(\mathbf{B}(Z''_{i+1}) \times \{p\}) = \Upsilon^i(X'')$. Nach Korollar 4.1 gilt $\Upsilon_{i+1} \in \text{Eiso}(\mathbf{B}(Z''_{i+1}) \times \{p\}, \overline{\mathbf{T}_{\mathcal{R}^{i+1}}}(Z''_{i+1}))$, womit

$$\begin{aligned} \Upsilon^{i+1}(X'') &= \Upsilon_{i+1}(\Upsilon^i(M \sqcup \text{Pa}(\mathbf{B}(Z''_{i+1}) \times \{p\}))) \\ &= \Upsilon_{i+1}(M) \sqcup \Upsilon_{i+1}(\text{Pa}_{\mathcal{R}^i}(\mathbf{B}(Z''_{i+1}) \times \{p\})) \\ &= M \sqcup \text{Pa}_{\mathcal{R}^{i+1}}(Z''_{i+1}) \end{aligned} \quad (9.7)$$

folgt.

Zu (9.7): Da, wie vorher gezeigt, $M \cap Z''_{i+1} = \emptyset$ gilt $\Upsilon_{i+1}(M) = M$.

Wir zeigen $\Upsilon_{i+1}(\text{Pa}(\mathbf{B}(Z''_{i+1}) \times \{p\})) = \text{Pa}_{\mathcal{R}^{i+1}}(Z''_{i+1})$. Es gilt

$$\mathcal{R}^{i+1} = \Upsilon_{i+1}(\overline{\mathbf{T}_{\mathcal{R}^i}(\mathbf{B}(Z''_{i+1}) \times \{p\})}) \cup \mathbf{T}_{\mathcal{R}}(Z''_{i+1}),$$

womit die einzigen Elemente, welche Z''_{i+1} als Teilmenge beinhalten, die Elemente der Kette $\Upsilon_{i+1}(\mathbf{K}_{\mathcal{R}^i}(\Upsilon^i(X''), \text{Pa}_{\mathcal{R}^i}(\mathbf{B}(Z''_{i+1}) \times \{p\})))$ sind. Da Υ_{i+1} , nach Korollar 4.1, ein Einbettungsisomorphismus ist, folgt, daß Υ_{i+1} ein Ordnungsisomorphismus ist. Damit ist das eindeutige minimale Element in \mathcal{R}^{i+1} , welches Z''_{i+1} enthält, $\Upsilon_{i+1}(\text{Pa}_{\mathcal{R}^i}(\mathbf{B}(Z''_{i+1}) \times \{p\}))$. Somit ist $\Upsilon_{i+1}(\text{Pa}_{\mathcal{R}^i}(\mathbf{B}(Z''_{i+1}) \times \{p\})) = \text{Pa}_{\mathcal{R}^{i+1}}(Z''_{i+1})$.

Aus $M \sqcup \text{Pa}_{\mathcal{R}^{i+1}}(Z''_{i+1}) = \Upsilon^{i+1}(X'')$ folgt abschließend

$$M \sqcup Z''_{i+1} = \Upsilon^{i+1}(X'').$$

□

Wir können nun den Einbettungsmonomorphismus φ_{neu} von \mathcal{G} nach \mathcal{R}^n konstruieren. Von Aussage 1 wissen wir, daß \mathcal{R}^n alle Teilbäume $T_{\mathcal{R}}(Z_i)$ mit $i \in \{1, \dots, n\}$ enthält. Daher behalten wir die partielle Abbildung $\varphi|_{T_{\mathcal{G}}(X) \setminus \{X\}}$ von φ für diese Elemente bei.

Die verbleibende Menge $\overline{T_{\mathcal{G}}}(X) \cup \{X\}$ wird durch $\varphi|_{\overline{T_{\mathcal{G}}}(X) \cup \{X\}}$ in genau den Urbildbereich von Υ^n abgebildet, womit $\Upsilon^n \circ \varphi|_{\overline{T_{\mathcal{G}}}(X) \cup \{X\}}$ wieder eine Funktion ist. Damit können wir φ_{neu} , wie folgt, als Funktion definieren.

Festlegung 2 $\varphi_{\text{neu}} : \mathcal{G} \rightarrow \mathcal{R}^n$ ist wie folgt definiert:

$$\varphi_{\text{neu}}(K) := \begin{cases} \Upsilon^n \circ \varphi|_{\overline{T_{\mathcal{G}}}(X) \cup \{X\}} & K \in \overline{T_{\mathcal{G}}}(X) \\ \varphi|_{T_{\mathcal{G}}(X) \setminus \{X\}} & \text{sonst.} \end{cases}$$

Aussage 3 $\varphi_{\text{neu}} \in \text{Emon}(\mathcal{G}, \mathcal{R}^n)$.

Beweis. Wir betrachten folgende Fälle:

1. $M, N \in \overline{T_{\mathcal{G}}}(X) \cup \{X\}$. Aus der Voraussetzung $\varphi \in \text{Emon}(\mathcal{G}, \mathcal{R})$ und der Abgeschlossenheit von $\overline{T_{\mathcal{G}}}(X) \cup \{X\}$ gegen \sqcup folgt

$$\varphi|_{\overline{T_{\mathcal{G}}}(X) \cup \{X\}} \in \text{Emon}(\overline{T_{\mathcal{G}}}(X) \cup \{X\}, \overline{T_{\mathcal{R}}}(X'') \cup \{X\}).$$

Nach Aussage 2 ist $\Upsilon^n \in \text{Emon}(\overline{T_{\mathcal{R}}}(X'') \cup \{X''\}, \overline{T_{\mathcal{R}^n}}(\Upsilon^n(X'')) \cup \Upsilon^n(X''))$. Somit ist $\Upsilon^n \circ \varphi|_{\overline{T_{\mathcal{G}}}(X) \cup \{X\}} \in \text{Emon}(\overline{T_{\mathcal{G}}}(X) \cup \{X\}, \overline{T_{\mathcal{R}^n}}(\Upsilon^n(X'')) \cup \Upsilon^n(X''))$, und es gilt $\varphi_{\text{neu}}(M) \sqcup \varphi_{\text{neu}}(N) = \varphi_{\text{neu}}(M \sqcup N)$.

2. $M, N \in T_{\mathcal{G}}(Z_i)$ mit $i \in \{1, \dots, n\}$. $T_{\mathcal{G}}(Z_i)$ ist gegen \sqcup abgeschlossen, und nach Voraussetzung gilt $\varphi \in \text{Emon}(\mathcal{G}, \mathcal{R})$.

Also ist $\varphi|_{T_{\mathcal{G}}(Z_i)} \in \text{Emon}(T_{\mathcal{G}}(Z_i), T_{\mathcal{R}}(Z_i''))$, und es gilt

$$\begin{aligned} \varphi_{\text{neu}}(M) \sqcup \varphi_{\text{neu}}(N) &= \varphi|_{T_{\mathcal{G}}(Z_i)}(M) \sqcup \varphi|_{T_{\mathcal{G}}(Z_i)}(N) \\ &= \varphi|_{T_{\mathcal{G}}(Z_i)}(M \sqcup N) = \varphi_{\text{neu}}(M \sqcup N). \end{aligned}$$

3. $M \in T_{\mathcal{G}}(Z_i)$, $N \in T_{\mathcal{G}}(Z_j)$ mit $i, j \in \{1, \dots, n\}$ und $i \neq j$. Nach Aussage 2 gilt $\varphi(Z_i) \sqcup \varphi(Z_j) = \Upsilon^n(\varphi(X))$, und durch Ersetzen mit φ_{neu} erhalten wir die äquivalente Formulierung $\varphi_{\text{neu}}(Z_i) \sqcup \varphi_{\text{neu}}(Z_j) = \varphi_{\text{neu}}(X)$.

Wir zeigen $\varphi_{\text{neu}}(X) = \varphi_{\text{neu}}(M \sqcup N)$. Da nach Voraussetzung $Z_i, Z_j \in \text{Ch}_{\mathcal{G}}(X)$ mit $i \neq j$ gilt, folgt $Z_i \cap Z_j = \emptyset$. Da nach Voraussetzung $M \subseteq Z_i$ und $N \subseteq Z_j$ gilt, folgt $M \sqcup N = Z_i \sqcup Z_j$. Aus $Z_i, Z_j \in \text{Ch}_{\mathcal{G}}$ folgt $Z_i \sqcup Z_j = X$. Also gilt $M \sqcup N = X$.

Wir zeigen $\varphi_{\text{neu}}(Z_i) \sqcup \varphi_{\text{neu}}(Z_j) = \varphi_{\text{neu}}(M) \sqcup \varphi_{\text{neu}}(N)$. Da nach Voraussetzung $\varphi \in \text{Emon}(\mathcal{G}, \mathcal{R})$ folgt $\varphi \in \text{Omon}(\mathcal{G}, \mathcal{R})$. Damit folgt $\varphi_{\text{neu}}(M) \subseteq \varphi_{\text{neu}}(Z_i)$ aus der Voraussetzung $M \subseteq Z_i$ und $\varphi_{\text{neu}}(N) \subseteq \varphi_{\text{neu}}(Z_j)$ aus der Voraussetzung $N \subseteq Z_j$. Weiter folgt $\varphi_{\text{neu}}(Z_i) \cap \varphi_{\text{neu}}(Z_j) = \emptyset$ aus $Z_i \cap Z_j = \emptyset$. Somit folgt aus $\varphi_{\text{neu}}(M) \subseteq \varphi_{\text{neu}}(Z_i)$, $\varphi_{\text{neu}}(N) \subseteq \varphi_{\text{neu}}(Z_j)$ und $\varphi_{\text{neu}}(Z_i) \subseteq \varphi_{\text{neu}}(Z_j)$, daß $\varphi_{\text{neu}}(N) \subseteq \varphi_{\text{neu}}(Z_i) \subseteq \varphi_{\text{neu}}(Z_j)$ gilt.

Abschließend folgt

$$\varphi_{\text{neu}}(M) \sqcup \varphi_{\text{neu}}(N) = \varphi_{\text{neu}}(Z_i) \sqcup \varphi_{\text{neu}}(Z_j) = \Upsilon^n(\varphi(X)) = \varphi_{\text{neu}}(M \sqcup N).$$

4. $M \in \mathcal{T}_{\mathcal{G}}(Z_i)$ mit $i \in \{1, \dots, n\}$, $N \in \overline{\mathcal{T}_{\mathcal{G}}}(X) \cup \{X\}$. Wir haben in Punkt 1 gezeigt, daß $\varphi_{\text{neu}}(X \sqcup N) = \varphi_{\text{neu}}(X) \sqcup \varphi_{\text{neu}}(N)$ gilt.

Wir zeigen $X \sqcup N = M \sqcup N$. Da $M \in \mathcal{T}_{\mathcal{G}}(Z_i)$ und $N \in \overline{\mathcal{T}_{\mathcal{G}}}(X) \cup \{X\}$, folgt $K_{\mathcal{G}}(Z_i, M) \cap K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), N) = \emptyset$. Damit gilt

$$\begin{aligned} X \sqcup N &= \min(K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), X) \cap K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), N)) \\ &= \min(K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), M) \cap K_{\mathcal{G}}(\text{Ro}(\mathcal{G}), N)) \\ &= M \sqcup N. \end{aligned}$$

Wir zeigen $\varphi_{\text{neu}}(X) \sqcup \varphi_{\text{neu}}(N) = \varphi_{\text{neu}}(M) \sqcup \varphi_{\text{neu}}(N)$. Es folgt $\varphi_{\text{neu}}(N) \in \overline{\mathcal{T}_{\mathcal{R}^n}}(\Upsilon^n(X''))$ aus $N \in \overline{\mathcal{T}_{\mathcal{G}}}(X) \cup \{X\}$. Mit $\Upsilon^n(X'') = \varphi_{\text{neu}}(X)$ folgt dann $\varphi_{\text{neu}}(N) \in \overline{\mathcal{T}_{\mathcal{R}^n}}(\varphi_{\text{neu}}(X))$. Aussage 2 besagt, $\Upsilon^n(Z_i'') \subset \Upsilon^n(X'')$, woraus $\varphi_{\text{neu}}(Z_i) \subset \varphi_{\text{neu}}(X)$ folgt. Damit folgt $K_{\mathcal{R}^n}(\text{Ro}(\mathcal{R}^n), \varphi_{\text{neu}}(N)) \cap K_{\mathcal{R}^n}(\varphi_{\text{neu}}(Z_i), \varphi_{\text{neu}}(M)) = \emptyset$, womit

$$\begin{aligned} \varphi_{\text{neu}}(X) \sqcup \varphi_{\text{neu}}(N) &= \min(K_{\mathcal{R}^n}(\text{Ro}(\mathcal{R}^n), \varphi_{\text{neu}}(X)) \cap K_{\mathcal{R}^n}(\text{Ro}(\mathcal{R}^n), \varphi_{\text{neu}}(N))) \\ &= \min((K_{\mathcal{R}^n}(\text{Ro}(\mathcal{R}^n), \varphi_{\text{neu}}(M)) \cap K_{\mathcal{R}^n}(\text{Ro}(\mathcal{R}^n), \varphi_{\text{neu}}(N))) \\ &= \varphi_{\text{neu}}(M) \sqcup \varphi_{\text{neu}}(N) \end{aligned}$$

gilt. Abschließend folgt

$$\varphi_{\text{neu}}(M) \sqcup \varphi_{\text{neu}}(N) = \varphi_{\text{neu}}(X \sqcup N) = \varphi_{\text{neu}}(X) \sqcup \varphi_{\text{neu}}(N) = \varphi_{\text{neu}}(M) \sqcup \varphi_{\text{neu}}(N).$$

□

Nach Lemma 4.5 gilt $B(K_{\mathcal{R}}(A, B)) = K_S(B(A), B(B))$ für alle $A, B \in \mathcal{R}^n$ mit $B \subseteq A$. Es bleibt zu zeigen, daß \mathcal{R}^n die Kardinalität von \mathcal{R} besitzt.

Aussage 4 \mathcal{R}^n besitzt minimale Kardinalität.

Beweis. Da \mathcal{R}^i für alle $i \in \{0, \dots, n\}$ nach Aussage 1 Ersetzungen sind, gilt:

$$|\mathcal{R}^0| = |\mathcal{R}| + |\mathbf{T}_s(\mathbf{B}(\varphi(\mathbf{X}))) \times p| - |\mathbf{T}_{\mathcal{R}}(\mathbf{X}'')|$$

$$|\mathcal{R}^i| = |\mathcal{R}^{i-1}| + |\mathbf{T}_{\mathcal{R}}(\mathbf{Z}_i'')| - |\mathbf{T}_{\mathcal{R}^{i-1}}(\mathbf{B}(\mathbf{Z}_i'') \times \{p\})| \text{ für alle } i \in \{1, \dots, n\}.$$

Weiter gilt

$$|\mathbf{T}_s(\mathbf{B}(\mathbf{X}'')) \times p| = |\mathbf{T}_s(\mathbf{B}(\mathbf{X}''))| \text{ und}$$

$$|\mathbf{T}_{\mathcal{R}^{i-1}}(\mathbf{B}(\mathbf{Z}_i'') \times \{p\})| = |\mathbf{T}_s(\mathbf{B}(\mathbf{Z}_i''))| \text{ für alle } i \in \{1, \dots, n\}.$$

Damit ist

$$\begin{aligned} |\mathcal{R}^n| &= |\mathcal{R}| - |\mathbf{T}_{\mathcal{R}}(\mathbf{X}'')| + |\mathbf{T}_s(\mathbf{B}(\mathbf{X}''))| \\ &\quad - \left(\sum_{i \in \{1, \dots, n\}} |\mathbf{T}_s(\mathbf{B}(\mathbf{Z}_i''))| \right) + \left(\sum_{i \in \{1, \dots, n\}} |\mathbf{T}_{\mathcal{R}}(\mathbf{Z}_i'')| \right) \\ &= |\mathcal{R}| - c \end{aligned}$$

mit

$$c = |\mathbf{T}_{\mathcal{R}}(\mathbf{X}'')| - |\mathbf{T}_s(\mathbf{B}(\mathbf{X}''))| + \left(\sum_{i \in \{1, \dots, n\}} |\mathbf{T}_s(\mathbf{B}(\mathbf{Z}_i''))| \right) - \left(\sum_{i \in \{1, \dots, n\}} |\mathbf{T}_{\mathcal{R}}(\mathbf{Z}_i'')| \right).$$

Aus Aussage 4.5 folgt

$$\mathbf{B}(\mathbf{T}_{\mathcal{R}}(\mathbf{X}'') \setminus \bigcup_{i \in \{1, \dots, n\}} \mathbf{T}_{\mathcal{R}}(\mathbf{Z}_i'')) = \mathbf{T}_s(\mathbf{B}(\mathbf{X}'')) \setminus \bigcup_{i \in \{1, \dots, n\}} \mathbf{T}_s(\mathbf{B}(\mathbf{Z}_i'')),$$

und damit gilt

$$|\mathbf{T}_{\mathcal{R}}(\mathbf{X}'')| - \sum_{i \in \{1, \dots, n\}} |\mathbf{T}_{\mathcal{R}}(\mathbf{Z}_i'')| \geq |\mathbf{T}_s(\mathbf{B}(\mathbf{X}''))| - \sum_{i \in \{1, \dots, n\}} |\mathbf{T}_s(\mathbf{B}(\mathbf{Z}_i''))|.$$

Also ist $c \geq 0$, und da \mathcal{R} nach Voraussetzung minimale Kardinalität besitzt, folgt $c = 0$, womit $|\mathcal{R}^n| = |\mathcal{R}|$ gilt. \square

Somit gilt $\mathcal{R}^n \in \text{RC}(\mathcal{G}, \mathcal{S})$. \square

Verw. von: Lem. 4.2, 4.5, Kor. 4.1; Verw. in: Lem. 4.5.

Literaturverzeichnis

- [Aho et al., 1976] Aho, A. V., Hopcroft, J. E., and Ullman, J. D. (1976). On finding lowest common ancestors in trees. *SIAM Journal on Computing*, 1(5):115–132.
- [Altschul et al., 1990] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Mathematical Biology*, 215:403–410.
- [Bairoch and Boeckmann, 1994] Bairoch, A. and Boeckmann, B. (1994). The SWISS-PROT protein sequence data bank: current status. *Nucleic Acids Research*, 22:3578–3580.
- [Bourque, 1978] Bourque, M. (1978). *Arbres de Steiner et reseaux dont varie l'emplacement de certains sommets*. PhD thesis, University de Montreal, Department d'Informatique et de Recherche Operationelle, University de Montreal, Quebec, Canada.
- [Cormen et al., 1996] Cormen, T. H., Leiserson, C. E., and Rivest, R. E. (1996). *Introduction to Algorithms*. MIT Press.
- [Estabrook and McMorris, 1980] Estabrook, G. F. and McMorris, F. R. (1980). When is one estimate of evolutionary relationships a refinement of another? *Journal of Mathematical Biology*, 10:367–373.
- [Estabrook et al., 1985] Estabrook, G. F., McMorris, F. R., and Meacham, C. A. (1985). Comparison of undirected trees based on subtrees of four evolutionary units. *Systematic Zoology*, 34:193–200.
- [Eulenstein et al., 1997] Eulenstein, O., Mirkin, B., and Vingron, M. (1997). Comparison of annotation duplication, tree mapping, and copying as methods to compare gene trees with species trees. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Sciences*, volume 37, pages 71–93.

- [Eulenstein et al., 1998] Eulenstein, O., Mirkin, B., and Vingron, M. (1998). Duplication-based measures of difference between gene- and species trees. *Journal of Computational Biology*, 5:135–148.
- [Eulenstein and Vingron, 1995] Eulenstein, O. and Vingron, M. (1995). On the equivalence of two tree mapping measures. Technical Report 936, GMD, Schloß Birlinghoven, 53754 St. Augustin, Germany.
- [Eulenstein and Vingron, 1998] Eulenstein, O. and Vingron, M. (1998). On the equivalence of two tree mapping measures. *Discr. Appl. Math.* in press.
- [Fitch, 1970] Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. *Systematic Zoology*, 19(2):99–113.
- [Goodman et al., 1979] Goodman, M., Czelusniak, J., Moore, G. W., Romero-Herrera, A. E., and Matsuda, G. (1979). Fitting the gene lineage into its species lineage: A parsimony strategy illustrated by cladograms constructed from globin sequences. *Systematic Zoology*, 28:132–168.
- [Guigó et al., 1996] Guigó, R., Muchnik, I., and Smith, T. F. (1996). Reconstruction of ancient molecular phylogeny. *Molecular Phylogenetics and Evolution*, 6(2):189–213.
- [Halmos, 1976] Halmos, P. R. (1976). *Naive Mengenlehre*. Vandenhoeck & Ruprecht, Göttingen, 4 edition.
- [Harel and Tarjan, 1984] Harel, D. and Tarjan, R. E. (1984). Fast algorithms for finding nearest common ancestors. *SIAM Journal on Computing*, 13(2):338–355.
- [Kawamura and Yokoyama, 1995] Kawamura, S. and Yokoyama, S. (1995). Paralogous origin of the rhodopsinlike opsin genes in lizards. *Journal of Molecular Evolution*, 40:594–600.
- [Li and Zhang, 1998] Li, M. M. and Zhang, L. (1998). On reconstructing species trees from gene trees in terms of duplications and losses. In *Proceedings of the 2nd International Conference on Computational Molecular Biology*.
- [Li and Graur, 1991] Li, W. and Graur, D. (1991). *Fundamentals of Molecular evolution*. Sinauer Associates, Inc., Sunderland, Massachusetts.

- [Mirkin et al., 1995] Mirkin, B., Muchnik, I., and Smith, T. F. (1995). A biology consistent model for comparing molecular phylogenies. *Journal of Computational Biology*, 2(4):493–507.
- [Nei, 1987] Nei, M. (1987). *Molecular Evolution Genetics*. Columbia University Press, New York.
- [Nelson and Platnik, 1981] Nelson, G. and Platnik, N. (1981). *Cladistics and vicariance*. Columbia University Press.
- [Okano et al., 1992] Okano, T., , Fukada, T. K. Y., Shijida, Y., Yoshizawa, T., and Page, R. D. M. (1992). Primary structures in chicken cone visual pigments. *Proceedings of the National Academy of Sciences*, 89:5932–5936.
- [Page, 1995] Page, R. (1995). COMPONENT. *Phylogenetic Program Package*, available at <http://taxonomy.zoology.gla.ac.uk/rod/cpw.html>.
- [Page, 1990a] Page, R. D. M. (1990a). Component analysis: A valiant failure? quantitative cladistic biogeography: Constructing and comparing area cladograms. *Cladistics*, 6(1):119–136.
- [Page, 1990b] Page, R. D. M. (1990b). Temporal congruence and cladistic analysis of biogeography and cospeciation. *Systematic Zoology*, 36:205–226.
- [Page, 1993] Page, R. D. M. (1993). Genes, organisms, and areas: The problem of multiple lineages. *Systematic Zoology*, 42(1):77–84.
- [Page, 1994a] Page, R. D. M. (1994a). Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Systematic Biology*, 43(1):58–77.
- [Page, 1994b] Page, R. D. M. (1994b). Parallel phylogenies: reconstructing the history of host-parasite assemblages. *Cladistics*, 10(1):155–173.
- [Page and Charleston, 1997] Page, R. D. M. and Charleston, M. A. (1997). Reconciled trees and incongruent gene and species trees. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Sciences*, volume 37.
- [Provencio et al., 1992] Provencio, I., Loew, E. R., and Forster, R. G. (1992). Vitamin A₂-based visual pigments in fully terrestrial vertebrates. *Vision Resolution*, 32:2201–2208.

- [Register et al., 1994] Register, E. A., Yokoyama, R., and Yokoyama, S. (1994). Multiple origins of the green-sensitive genes in fish. *Journal of Molecular Evolution*, 39:268–273.
- [Robinson, 1971] Robinson, D. F. (1971). Comparison of labeled trees with valency three. *Journal of Combinatorial Theory*, 11:105–119.
- [Robinson and Foulds, 1981] Robinson, D. F. and Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, 53:131–147.
- [Rosen, 1978] Rosen, D. E. (1978). Vicariant patterns and historical explanation biogeography. *Systematic Zoology*, 27:159–188.
- [Saitou and Nei, 1987] Saitou, N. and Nei, N. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Journal of Molecular Biology and Evolution*, 4:406–425.
- [Slowinski et al., 1997] Slowinski, J. B., Knight, A., and Rooney, A. P. (1997). Inferring species trees from gene trees: A phylogenetic analysis of the elapidæ (serpentes) based on the amino acid sequences of venom proteins. *Molecular Phylogenetics and Evolution*, 8(3):349–362.
- [Stryer, 1995] Stryer, L. (1995). *Biochemistry*. W. H. Freeman and Company.
- [Tarjan, 1979] Tarjan, R. E. (1979). A class of algorithms which require non-linear time to maintain disjoint sets. *Journal of Computer and System Sciences*, 18:110–127.
- [Tatusov et al., 1997] Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997). A genomic perspective on protein families. *Science*, 278:631–637.
- [Thompson et al., 1994] Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673–4680.
- [Wang et al., 1992] Wang, S.-Z., Adler, R., and Nathans, J. (1992). A visual pigment from chicken that resembles rhopsin: amino acid sequence, gene structure, and functional expression. *Biochemistry*, 31:3309–3315.
- [Waterman and Smith, 1978] Waterman, M. S. and Smith, T. S. (1978). On the similarity of dendrograms. *Journal of Theoretical Biology*, 73:789–800.

- [Yuan et al., 1998] Yuan, Y. P., Eulenstein, O., Vingron, M., and Bork, P. (1998). Towards detection of orthologues in sequence databases. *Bioinformatics*. in press.
- [Zhang, 1997] Zhang, L. (1997). On a mirkin-muchnik-smith conjecture for comparing molecular phylogenies. *Journal of Computational Biology*, 4(2):177-187.