

Biodiversitätsmessung bei Pflanzen anhand molekularer Daten: Ein Beitrag zur wissenschaftlichen Definition von Biodiversität

Jan T. Kim, Henning Schwöbbermeyer, Günter Theißen und Heinz Saedler
Max-Planck-Institut für Züchtungsforschung

Inhaltsverzeichnis

1 Zielsetzung	3
1.1 Motivation der Zielsetzung	3
2 Biologische Grundlagen	3
2.1 Genetische Information	3
2.2 Entwicklung	4
2.3 Mutation	4
2.4 DNA-Reparatur	5
2.5 Transposons	5
2.6 Evolution	6
3 Problemstellung / Stand der Erkenntnis	6
3.1 Probleme des Artkonzeptes	6
3.2 Verschiedene Ebenen der Biodiversität	7
3.3 Arbeitsansatz	8
4 Biodiversität auf genetischer und morphologischer Ebene	8
4.1 Biodiversität entsteht durch Evolution	8
4.2 Transposons sind Generatoren von Biodiversität	8
4.2.1 Klassifikation von Transposons und Transpositionsmechanismen	9
4.2.2 Erzeugung genetischer Diversität durch transposonbedingte Rearrangements	10
4.2.3 Transposonbasierte Entstehung phänotypischer Diversität	11
4.2.4 Die Bedeutung von Transposons für Evolution und Biodiversität	12
4.3 Biodiversität auf genetischer und morphologischer Ebene am Beispiel von Pflanzen	13
4.3.1 Drastische Unterschiede in der Diversität auf molekulargenetischer und morphologischer Ebene: das Fallbeispiel Teosinte-Mais	13
4.3.2 Vom Genom zum Phänotyp	15
4.4 Schlußfolgerungen	18
5 Ein Ansatz zur quantitativen Charakterisierung von Biodiversität	19
5.1 Distanzen	19
5.2 Diversitätsmaße auf der Grundlage von Distanzen	19
5.2.1 Die Biodiversitätsmaße von WILLIAMS, HUMPHRIES und VANE-WRIGHT	20
5.2.2 Formale Kriterien für Biodiversitätsmaße	21
5.2.3 Die Baumlänge als Biodiversitätsmaß	21
5.2.4 Stammbaunabhängige distanzbasierte Biodiversitätsmessung	21
5.2.5 Weitere Biodiversitätsmaße	22
5.3 Die Distanzverteilungskomplexität (DVK) als Biodiversitätsmaß	22
5.3.1 Definition der Distanzverteilungskomplexität	22
5.3.2 Diskretisierung bei realwertigen Distanzmaßen	23

6	Untersuchung und Vergleich verschiedener Biodiversitätsmaße auf der Grundlage molekularer Daten	23
6.1	Material und Methoden	23
6.1.1	Eingabedaten	23
6.1.2	Biodiversitätsmaße	24
6.1.3	Wurzelpositionsverschiebung	24
6.1.4	Simulation von Biodiversitätsverlusten	24
6.2	Ergebnisse und Diskussion	25
6.2.1	DVK-Berechnung mit unterschiedlichen Diskretisierungsintervallgrößen	25
6.2.2	Abhängigkeit von der Wurzelposition	27
6.2.3	Charakterisierung von Biodiversitätsverlusten	27
6.3	Zusammenfassende Diskussion	32
7	Zusammenfassung	33
	Literatur	34

1 Zielsetzung

Ziel unserer Arbeit ist es, einen Beitrag zur wissenschaftlichen Definition des Begriffes „Biodiversität“ zu leisten. Dazu vergleichen wir Verfahren zur Biodiversitätsmessung anhand molekulargenetischer Daten aus Pflanzen auf ihre Eignung, Biodiversität quantitativ zu charakterisieren. Wir beschränken uns dabei auf solche Verfahren, die auf allen Ebenen biologischer Organisation anwendbar sind, auf denen Biodiversität beschrieben wird [EHRlich und WILSON 1991, SOULÉ 1991, WILSON 1993].

Grundlage dieses Beitrags ist die Annahme, daß Biodiversität ein Bestandteil biologischer Realität ist, welcher wissenschaftlich untersucht und beschrieben werden kann. Zur Untersuchung der Biodiversität können – und sollen – verschiedene Methoden, die unterschiedlichen biologischen Disziplinen zugerechnet werden können, eingesetzt werden. Wir widersprechen jedoch der im Kapitel 8 dargestellten Auffassung, nach der durch die unterschiedlichen Untersuchungsmethoden verschiedene Forschungsgegenstände konstituiert werden. Vielmehr gehen wir davon aus, daß Begriffe wie „Gen“ und „Biodiversität“ reale Gegenstände bezeichnen, deren Identität unabhängig davon ist, mit welchen Methoden sie untersucht werden. PUTNAM [PUTNAM 1993] hat dies für das Gen so formuliert:

Sicherlich ist das „Gen“, was in der Molekularbiologie behandelt wird, eben jenes Gen (oder jener „Faktor“), von dem Mendel sprechen wollte – zweifellos ist es das, wovon zu sprechen er die Absicht gehabt haben sollte!

Wir möchten in diesem Sinne mit unserem Beitrag *einen Forschungsgegenstand*, nämlich die Biodiversität, mit *verschiedenen Methoden* untersuchen, ohne dadurch die Existenz verschiedener Biodiversitäten zu behaupten.

1.1 Motivation der Zielsetzung

Biodiversität ist ein Begriff, der in verschiedenen wissenschaftlichen Fachkreisen verwendet wird, aber auch weit darüber hinaus (etwa in der Politik). Selbst in der Wissenschaft fehlt jedoch bislang eine präzise Definition dieses Begriffs, die in allen beteiligten Fachkreisen in sinnvoller Weise anwendbar ist und die somit für alle Wissenschaftler, die im interdisziplinären Forschungsfeld der Biodiversität aktiv sind, als gemeinsame Arbeitsgrundlage dienen kann. Infolge dieses Mangels existieren in verschiedenen Fachkreisen und Teilen der Öffentlichkeit verschiedene, teils voneinander stark abweichende Vorstellungen darüber, was Biodiversität eigentlich ist. In wissenschaftlichen Arbeiten werden zur Quantifizierung von Biodiversität derzeit verschiedene Indikatoren, wie etwa Artenreichtum, verwendet. Die Verwendung dieser Indikatoren erfolgt aufgrund unterschiedlicher, z.T. inkompatibler Biodiversitätskonzepte. Das Fehlen eines einheitlichen, eindeutig definierten Biodiversitätskonzepts ist für die wissenschaftliche Forschung eine unbefriedigende Situation. In der öffentlichen Diskussion führt es zu Schwierigkeiten bei der Kommunikation und resultiert in vielerlei Mißverständnissen. Bei manchen politischen Entscheidungsträgern ist bereits der Eindruck entstanden, daß zum Themenbereich der Biodiversität keine kompetente wissenschaftliche Beratung zu erwarten sei. Mit unseren Ansätzen zur Definition des Biodiversitätsbegriffs möchten wir einen Beitrag zur wissenschaftlichen Untersuchung des Phänomens Biodiversität und zur Entwicklung einer Grundlage für die sachliche Diskussion über Biodiversität in der Öffentlichkeit leisten.

2 Biologische Grundlagen

Die folgenden Abschnitte sollen lediglich einen kurzen Abriß der biologischen Grundlagen unserer Ausführungen geben. Ausführliche Informationen zu den hier umrissenen Themenfeldern finden sich in Standardwerken der Genetik (z.B. [HAGEMANN 1999, KNIPPERS 1997]).

2.1 Genetische Information

DNA (*desoxyribonucleic acid*) ist ein polymeres Molekül, das aus den Nucleotidkomponenten desoxy-Adenosin (A), -Cytosin (C), -Guanosin (G) und -Thymidin (T) besteht. Durch die Sequenz dieser Nucleotide in DNA-Molekülen ist die Erbinformation aller Lebewesen codiert. Die Verknüpfung der Nucleotide erfolgt jeweils über Phosphodiesterbrücken, die das 3'-Kohlenstoffatom im Desoxyribosebestandteil eines Nucleotids und das 5'-Kohlenstoffatom eines benachbarten Nucleotids verbinden. Durch diese Asymmetrie hat die Nucleotidsequenz eines DNA-Strangs eine definierbare Polarität. Per Konvention werden Sequenzen durch Zeichenketten notiert,

bei denen das Ende mit einem freien 5'-Kohlenstoffatom links und das freie 3'-Ende dementsprechend rechts geschrieben wird.

G und C sowie A und T können sogenannte Watson-Crick-Basenpaare bilden und werden daher als komplementär bezeichnet. Außer bei bestimmten Viren liegt DNA stets in der Form antiparalleler, komplementärer Doppelstränge vor. Die Basenkomplementarität ist auch die Grundlage der Replikation von DNA. Bei dieser wird ein Doppelstrang aufgetrennt und zu beiden Einzelsträngen wird anschließend jeweils ein neuer, komplementärer Strang synthetisiert. Das Ergebnis dieses Prozesses sind zwei identische Doppelstränge.

DNA-Sequenzen lassen sich ihrer biologischen Funktion entsprechend in Gene einteilen. Die Decodierung der genetischen Information erfolgt durch Interaktion mit verschiedenen molekularen Komponenten einer Zelle. Wesentliche Vorgänge dabei sind:

- Die Aktivierung von Genen durch Transkriptionsfaktoren
- Die Transkription von DNA-Sequenzen in RNA-Sequenzen durch DNA-abhängige RNA-Polymerasen; je nach Gen werden hierbei für Proteine codierende *messenger*-RNAs, ribosomale RNAs, tRNAs usw. synthetisiert
- Die Translation von *messenger*-RNAs in Proteine, die auf der molekularen Ebene den größten Teil biologischer Funktion realisieren.

Der Promotor eines Gens entscheidet weitgehend über die Aktivierung des Gens. In der Regel liegt der Promotor stromaufwärts des Transkriptionsstarts. Er enthält Sequenzelemente, an welche Transkriptionsfaktoren in spezifischer Weise binden können. Die Sequenzelemente des Promotors werden *cis*-regulatorische DNA-Elemente genannt, die Transkriptionsfaktoren werden auch als *trans*-regulatorische Proteine bezeichnet. Die Bindung *trans*-regulatorischer Proteine an *cis*-regulatorische Elemente kann die Transkription des betreffenden Gens sowohl aktivieren als auch inhibieren.

Der Aufbau sämtlicher für die Decodierung der genetischen Information erforderlichen molekularen Komponenten ist in genetischer Information codiert. Eine Zelle umfaßt also einen vollständigen Satz genetischer Information und einen vollständigen Mechanismus zu deren Decodierung.

2.2 Entwicklung

Die Entwicklung aller Lebewesen wird durch genetische Information gesteuert. Erkenntnisse der Molekulargenetik haben insbesondere bei eukaryotischen, vielzelligen Lebensformen ergeben, daß das komplexe Zusammenspiel einer Vielzahl von Transkriptionsfaktoren bei der Umsetzung der genetischen Information in morphologische Strukturen eine zentrale Rolle spielt. Ergebnis dieses Zusammenspiels ist ein spatiotemporales Muster von Genaktivitäten, die in ihrer Gesamtheit in der Realisierung der Ontogenese eines Organismus resultieren. Da höhere, vielzellige Organismen, wie Tiere und Pflanzen, besonders komplex sind, wird zur Steuerung ihrer Entwicklung besonders viel genetische Information benötigt.

2.3 Mutation

Die Sequenz von DNA-Molekülen kann durch eine Vielzahl von Prozessen temporär oder erblich verändert werden. Erbliche Veränderungen werden als Mutationen bezeichnet. Hierbei ist zu unterscheiden zwischen elementaren Mutationen und Sequenzrearrangements. Elementare Mutationen sind Punktmutationen, bei denen nur ein Basenpaar verändert wird, und das Einfügen sowie Entfernen einzelner Basenpaare. Die Rearrangements lassen sich weiter in Deletionen (Entfernung eines Sequenzabschnitts), Inversionen (Umkehr der Orientierung eines Sequenzabschnitts), Duplikationen (Verdopplung eines Sequenzabschnitts), Translokationen (Verlagerung eines Sequenzabschnitts) sowie Transposition (s. Abschnitt 2.5) einteilen.

Alle Mutationen verändern die genetische Information in ungerichteter Weise. Die biologische Funktion eines Gens, also die Struktur und Funktion der von ihm codierten Biomoleküle sowie sein Expressionsmuster, wird durch Mutationen nicht notwendigerweise verändert. Mutationen, die die biologische Funktion nicht verändern, haben dementsprechend keinen Einfluß auf die Überlebens- und Fortpflanzungschancen eines betroffenen Individuums. Daher werden solche Mutationen als neutral (gemeint ist dies im Hinblick auf den Selektionswert, s. auch Abschnitt 2.6) bezeichnet.



Abbildung 1: Von Transposons verursachte Mosaikfarbmuster (a) in der Blüte des Löwenmäulchens, (b) auf Maiskörnern

2.4 DNA-Reparatur

Die Mehrheit der nichtneutralen Mutationen ist für die betroffenen Individuen nachteilhaft. Daraus ergibt sich ein Selektionsdruck für die Verringerung der Mutationsraten. Unter diesem Druck entstand im Laufe der Evolution eine Vielzahl molekularer Mechanismen, die die Genauigkeit der Replikation verbessern (*proofreading*), die die DNA vor schädlichen Einflüssen schützen (z.B. Hautbräune zum Schutz vor Sonnenstrahlung), und die beschädigte DNA erkennen und reparieren.

2.5 Transposons

Die Anordnung von Basenpaaren und Genen in der DNA bleibt im Regelfall konstant. Durch diese Stabilität der Genpositionen auf den Chromosomen wird der Austausch genetischer Information bei der geschlechtlichen Fortpflanzung erst möglich. Es gibt jedoch eine Klasse von DNA-Elementen, die von einer chromosomalen Position zu einer anderen wandern können. Dieser Vorgang wird als Transposition bezeichnet, und die entsprechenden Elemente werden Transposons (oder auch „springende Gene“) genannt.

Transpositionen sind eine spezielle Form von Mutation, die eine Vielzahl biologischer Effekte haben können. Gene, in die ein Transposon „hineinspringt“, werden durch dieses Ereignis in vielen Fällen inaktiviert. Es ist jedoch auch möglich, daß das Expressionsmuster des betroffenen Gens verändert wird. Darüberhinaus bleiben bei der Excision von Transposons oftmals kleinere Mutationen zurück. Insbesondere bei den in Pflanzen vorkommenden Transposons verbleibt nach der Excision oft ein *footprint*, ein Sequenzstück mit einer Länge von wenigen Basen, das oft eine charakteristische innere Symmetrie aufweist.

Bei der Excision eines Transposons wird in vielen Fällen die ursprüngliche Funktion des betroffenen Gens wiederhergestellt. Die entsprechenden durch die Integration eines Transposon bedingten Mutationen werden daher instabil genannt. Excisionen und die damit einhergehenden phänotypischen Reversionen können mit hoher Rate bei somatischen Zellteilungen während der Ontogenese eintreten. In solchen Fällen entwickelt sich ein Organismus, in dem Gruppen oder Sektoren von Zellen das Gen in seiner funktionalen Form enthalten.

Bei Genen, die eine Rolle bei der Synthese von Farbpigmenten (z.B. in der Blüte oder im Maiskorn, Abb. 1) spielen, kann die Instabilität und die resultierende Mosaikbildung anhand der Ausprägung der entsprechenden Far-

be verfolgt werden, weil alle Nachfahren einer Zelle, bei der das Transposon den entsprechenden Locus verlassen hat, das Pigment bilden und einen sichtbaren farbigen Fleck oder Sektor bilden. Anhand dieses Phänomens kann die in manchen Fällen extrem hohe Excisionsrate unmittelbar beobachtet werden, die ein geflecktes Mosaikmuster zur Folge hat („Variegation“). Detailliertere Ausführungen zu den Mechanismen der Transposition und den biologischen Bedeutung von Transposons finden sich in Abschnitt 4.2.

2.6 Evolution

Mutationen, Transpositionen usw. erzeugen in jeder Population ständig neue Variationen von DNA-Sequenzen. Ein erheblicher Teil dieser Vielfalt bleibt jedoch nicht für längere Zeit erhalten. Hierfür sind verschiedene Mechanismen verantwortlich, die zusammenfassend als Selektion bezeichnet werden:

Nichtneutrale Selektion: Bei nicht neutralen Mutationen sind die mutanten Formen in den meisten Fällen hinsichtlich ihrer Überlebens- und Fortpflanzungschancen dem Wildtyp gegenüber benachteiligt und werden aus den Populationen verdrängt. Wenn Mutanten auftreten, die hinsichtlich ihrer Überlebens- und Fortpflanzungschancen dem Wildtyp überlegen sind, verdrängen diese Mutanten den bisherigen Wildtyp.

Neutrale Selektion: Bei Mutationen, die annähernd oder vollständig neutral sind, d.h. bei denen sich die Mutanten phänotypisch vom Wildtyp kaum oder überhaupt nicht unterscheiden, kommt es zu zufälligen Fluktuationen der Häufigkeiten von Wildtyp- und Mutantallel. Im Verlauf langer Zeiträume sinkt dabei die Häufigkeit eines Allels auf Null ab, das entsprechende Allel geht dann verloren.

Die natürliche Selektion, insbesondere die nichtneutrale Selektion, ist also ein biologischer Prozeß, der die durch Mutationen erzeugte, ungerichtete genetische Variation strukturiert. Die Evolution, die durch das Zusammenwirken von Mutation und Selektion zustandekommt, kann somit als Mechanismus, der eine strukturierte genetische Diversität hervorbringt, charakterisiert werden.

Bei der Selektion ist zu beachten, daß die Reproduktionschance eines Individuums in wesentlicher Weise durch andere Organismen, mit denen das Individuum interagiert, bestimmt wird. Koevolutionäre Mechanismen spielen also eine entscheidende Rolle beim Evolutionsgeschehen; die „Fitness“ eines Organismus kann daher nur im Zusammenhang mit seinem ökosystemaren Kontext, der seinerseits evolutionärer Veränderung unterliegt, bestimmt werden. Durch die evolutionäre Veränderung der biotischen Umweltfaktoren entstehen neue ökologische Nischen, die die Evolution neuer Arten gestatten. Infolge dieser koevolutionären Dynamik kommt es nicht zur Konvergenz der evolutionären Adaptationsprozesse (wie dies regelmäßig für evolutionäre Algorithmen beschrieben wird), sondern genetische Information und Diversität sowie Biodiversität auf allen anderen Ebenen biologischer Organisation entwickeln sich in einem offenen Evolutionsprozeß (*open ended evolution*) fortlaufend weiter.

3 Problemstellung / Stand der Erkenntnis

Weshalb besteht Bedarf für eine naturwissenschaftliche Definition des Begriffs „Biodiversität“?

Der Begriff „Biodiversität“ wird vielfältig verwendet, von der Grundlagenforschung bis zur Philosophie. Darüberhinaus ist er von erheblicher politischer Brisanz, insbesondere wenn es um die Fragen geht, wieviel Biodiversität durch menschliche Einflüsse verloren geht, wer dafür verantwortlich zu machen ist und welche Gegenmaßnahmen ergriffen werden sollen.

Als ein aus den Naturwissenschaften stammender Terminus sollte die Verwendung des Begriffes Biodiversität auf einer soliden definitorischen Grundlage erfolgen, damit alle wissen, wovon die Rede ist. Was also ist Biodiversität? Leider ist bislang darauf keine befriedigende Antwort gegeben worden. Dennoch bedient man sich verschiedener Hilfsindikatoren, deren wichtigster wahrscheinlich der Begriff „Artenreichtum“ ist. „Artenreichtum“ ist aber aus mehreren Gründen nur ein unzureichender Ersatz für den Begriff „Biodiversität“, wie nachfolgend dargelegt wird.

3.1 Probleme des Artkonzeptes

Bisherige Arbeiten zur Biodiversität setzen Biodiversität zwar oft nicht gleich mit Artenreichtum, verwenden Artenreichtum aber als Indikator in Ermangelung einer exakten Definition von Biodiversität (siehe z.B. [McGRADY-STEED et al. 1997]). Doch dies ist aus mehreren Gründen problematisch.

Zunächst muß festgestellt werden, daß es in der Biologie kein einheitliches Artkonzept gibt, vielmehr ist der Artbegriff selbst in der Biologie äußerst umstritten. Es sei hier nur auf so unterschiedliche Konzepte wie das „biological species concept“ (BSC) und das „phylogenetic species concept“ (PSC) verwiesen (für einen Überblick über den Stand der Diskussion, siehe [AVISE und WOLLENBERG 1997, BACHMANN 1998]). Dies führt bei allen Biodiversitätsmaßen, die auf ein Spezieskonzept zurückgreifen, zu dem Problem, daß die mit den Maßen erhaltenen Ergebnisse u.U. in kritischer Weise von der Wahl des Spezieskonzepts abhängen (s. z.B. [BARRATT et al. 1997, BACHMANN 1998]).

Eine weitere Einschränkung resultiert daraus, daß keines der derzeitigen Spezieskonzepte für alle Lebensformen anwendbar ist. Dies begrenzt die Möglichkeiten, die Entwicklung von Biodiversität in verschiedenen Reichen und Domänen des Lebens anhand von artbasierten Biodiversitätsmaßen vergleichend zu untersuchen.

Selbst wenn ein einheitliches Spezieskonzept verfügbar wäre, müßte man mit systematischen Verzerrungen rechnen, weil verschiedene Taxa sich, insbesondere auf den höheren taxonomischen Kategorien, in ihrer biologischen Dynamik erheblich unterscheiden. Zum Beispiel definiert das biologische Spezieskonzept eine Art als eine Menge von Individuen, die sich miteinander fruchtbar kreuzen können. Die durch diese Definition implizierten Artgrenzen sind jedoch im Tierreich erheblich schärfer ausgebildet als bei den Pflanzen. Die Dynamik der Artentstehung im Sinne des biologischen Spezieskonzepts ist also im Tierreich und im Pflanzenreich unterschiedlich. Biodiversitätswerte, die auf der Grundlage des biologischen Spezieskonzepts ermittelt wurden, sind daher nur sehr eingeschränkt vergleichbar und man muß mit einer Unterschätzung der Biodiversität im Pflanzenreich, also mit einer systematischen Verzerrung, rechnen.

Bei der Verwendung der Anzahl verschiedener Arten als Biodiversitätsmaß ergibt sich - neben der kritischen Abhängigkeit von der Wahl des zugrundegelegten Artkonzepts - ein weiteres Problem. Bei der schlichten Erfassung der Artenzahl wird die Verschiedenheit der Arten nicht berücksichtigt. Die Biodiversität in einem Terrarium mit 1000 verschiedenen Insektenarten wäre, wenn man nur die Artenzahl betrachtet, ebenso groß wie in einem Terrarium von 1000 Arten von Pflanzen, Mikroben und unterschiedlichen Tieren, die gemeinsam ein Ökosystem bilden. Ein Ansatz zur Vermeidung dieses Problems ist der Übergang zu höheren taxonomischen Kategorien. So ließe sich argumentieren, daß tausend Insektenarten nur eine Kategorie von Lebewesen repräsentieren, während ein Ökosystem von Pflanzen, Mikroben und Tieren eine größere Anzahl von Kategorien enthält. Es lassen sich jedoch, analog zu obigem Beispiel, Fälle konstruieren, bei denen die Zählung der Taxa auf jeder beliebigen Ebene zu unsinnigen Ergebnissen führt. Zudem verliert man bei der Zählung von Taxa auf höheren Ebenen die Information über die Diversität innerhalb der jeweiligen Taxa sowie auf allen Ebenen unterhalb der Taxa (z.B. intraspezifische Diversität bei Zählung auf Artebene). Wollte man zum Beispiel die tierische Biodiversität auf der Stammesebene bestimmen, so gelangte man zu dem Befund, daß die Biodiversität des Meeres die der terrestrischen Ökosysteme bei weitem überträfe. Der Grund hierfür besteht darin, daß während der Evolution des Lebens von den ca. 35 Stämmen des Tierreiches nur 7 den „Sprung aufs Land“ geschafft haben [RAFF 1996]. Dies steht in direktem Widerspruch dazu, daß terrestrische Ökosysteme, vor allem tropische Regenwälder, üblicherweise als Zentren der Biodiversität angesehen werden.

Erschwerend kommt bei derart simplen Definitionen von Biodiversität hinzu, daß Arten- oder Bauplanreichtum in einem bestimmten Areal nicht die Häufigkeit des Auftretens der verschiedenen Formen berücksichtigen. So könnte ein bestimmter Bauplan nur durch ein letztes Restexemplar einer einzigen Spezies repräsentiert sein und damit eine größere Diversität vorspiegeln, als im Biotop nach jeder sinnvollen Definition vorhanden ist.

3.2 Verschiedene Ebenen der Biodiversität

Im vorhergegangenen Abschnitt klang auch ein weiteres Problem der Einschränkung von Biodiversität auf die Artebene an: eine derartige Einschränkung wird nicht der Tatsache gerecht, daß biologische Vielfalt auf vielen Ebenen existiert, von der molekularen und genomischen Ebene über die der Organismen (Individuen) und Arten, bis zur Ebene der Ökosysteme. Besonders faszinierend ist hierbei die Frage, inwieweit Diversität auf den verschiedenen Ebenen kausal miteinander verknüpft ist. Dies leitet zu dem viel diskutierten Problem über, wie der Genotyp eines Organismus seinen Phänotyp bestimmt und dieser wiederum die ökologische Potenz eines Lebewesens beeinflusst. Dies soll im Folgenden exemplarisch anhand der Beziehung zwischen Genotyp (auf molekularer Ebene) und Phänotyp bei grünen Pflanzen erläutert werden.

3.3 Arbeitsansatz

Alle Lebensformen auf der Erde besitzen Erbinformation, die in Basensequenzen von Nucleinsäuremolekülen (s. Abschnitt 2.1) codiert ist. Ferner sind bei allen Lebewesen der Mechanismus der Replikation von Erbinformation sowie der Mechanismus der Translation von Nucleotidsequenzen in Aminosäuresequenzen in allen wesentlichen Belangen gleich. Der Besitz genetischer Information sowie die fundamentalen Mechanismen ihrer Replikation und ihrer Decodierung sind also universelle Eigenschaften lebender Systeme.

Die genetische Information bietet sich deshalb als Grundlage für die Bestimmung von Biodiversität an. Aufgrund ihrer Universalität gestattet sie die einheitliche Einbeziehung von Lebensformen aus allen Reichen und Domänen des Lebens in die Charakterisierung von Biodiversität. Anders als bei einer spezieorientierten Biodiversitätsbestimmung sind hier also keine unterschiedlichen Konzepte für verschiedene Lebensformen erforderlich.

Eine weitere, wesentliche Eigenschaft der genetischen Information ist ihre digitale Struktur. Neben und zwischen den vier Basen gibt es keine Zwischenformen. Sequenzen können daher, anders als die Zuordnung von Organismen zu Arten, ohne Ermessensspielräume und willkürliche Entscheidungen bestimmt werden.

Wir wählen aus diesen Gründen die Ebene genetische Information als Grundlage für die Bestimmung von Biodiversität. Im folgenden Abschnitt 4 beschreiben wir zunächst Schlüsselaspekte der Dynamik genetischer Information und der Umsetzung genetischer Information in morphologische Strukturen. In den Abschnitten 5 und 6 untersuchen wir Maße, mit denen anhand genetischer Information Biodiversität charakterisiert werden kann. Die Anwendbarkeit dieser Maße ist jedoch nicht auf die genetische Ebene beschränkt. Da Biodiversität auf vielen Ebenen biologischer Organisation auftritt, sollte sie auch auf entsprechend vielen Ebenen charakterisiert werden können.

4 Biodiversität auf genetischer und morphologischer Ebene

4.1 Biodiversität entsteht durch Evolution

Evolution (s. Abschnitt 2.6) ist der biologische Mechanismus, der Biodiversität erzeugt [HANSON et al. 1999]. Das Grundprinzip aller Evolutionsprozesse ist ein Zusammenspiel von Quellen ungerichteter Diversität mit Selektionsmechanismen, welche diese Diversität fortlaufend strukturieren. Bei der biotischen Evolution sind Mutationen die Quelle ungerichteter Diversität. Die biologischen Selektionsmechanismen sind komplex; und verschiedene Mechanismen können ihre Wirkung auf unterschiedlichen Ebenen der Realisierung phänotypischer Strukturen entfalten.

Zu den von der Evolution hervorgebrachten Strukturen gehören auch die DNA-Polymerasen mit *proofreading*-Funktion, verschiedene Enzymsysteme zur Reparatur beschädigter DNA (s. Abschnitt 2.3) sowie die Enzymsysteme für die Transposition (s. Abschnitt 2.5). Die Frequenz von Mutationsvorgängen unterschiedlicher Art ist also ein Gegenstand evolutionärer Entwicklung, wobei Mutationsarten, welche die biologische Bedeutung der genetischen Information stärker entstellen, im Laufe der Evolution zunehmend reprimiert werden. Diese evolutionäre Strukturierung der Mutationsprozesse wirkt sich auch auf die Struktur der Biodiversität aus.

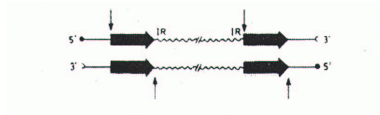
Die morphologische Struktur eines Lebewesens ist für seine Überlebens- und Fortpflanzungsfähigkeit von entscheidender Bedeutung. Die Morphologie ist damit eine phänotypische Ebene, auf der einige der wichtigsten Selektionsmechanismen ansetzen. Die Umsetzung genetischer Information in phänotypische Strukturen ist daher für die Evolution, und insbesondere für die Entstehung von Biodiversität, von entscheidender Bedeutung.

Die evolutionäre Gestaltung der Mutationsmechanismen und die Genotyp-Phänotyp-Beziehung sind faszinierende und komplexe Gegenstände biologischer Forschung, und eine systematische Zusammenfassung des Erkenntnisstandes auf diesen Feldern würde den Rahmen des vorliegenden Beitrags bei weitem überschreiten. In den folgenden Abschnitten werden diese Schlüsselaspekte der Evolution von Biodiversität daher jeweils anhand eines ausgewählten Beispiels exemplarisch erläutert.

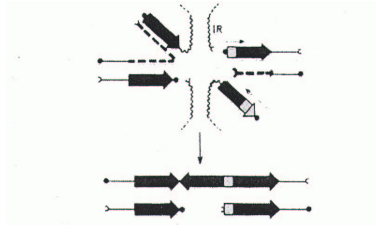
4.2 Transposons sind Generatoren von Biodiversität

Transposons, oder mobile genetische Elemente, kommen in Pflanzen, Tieren und Bakterien vor. Sie wurden von Barbara McClintock 1951 in Mais entdeckt und haben sich insbesondere bei Pflanzen als bedeutende Faktoren bei der Erzeugung genetischer Diversität erwiesen. Die definierende Eigenschaft von Transposons ist ihre Mobilität innerhalb des Genoms.

Eine Transposase erzeugt versetzte Schnitte



Hypothetische Intermediate



Mögliche Excisionsprodukte

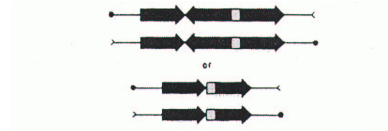


Abbildung 2: Das Exonucleasemodell der Excision pflanzlicher Klasse II Transposons

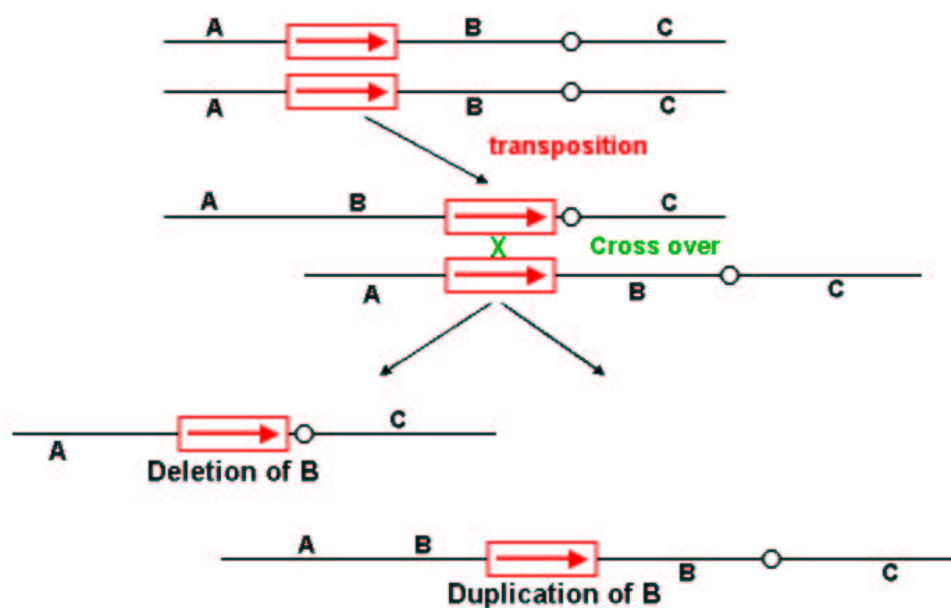
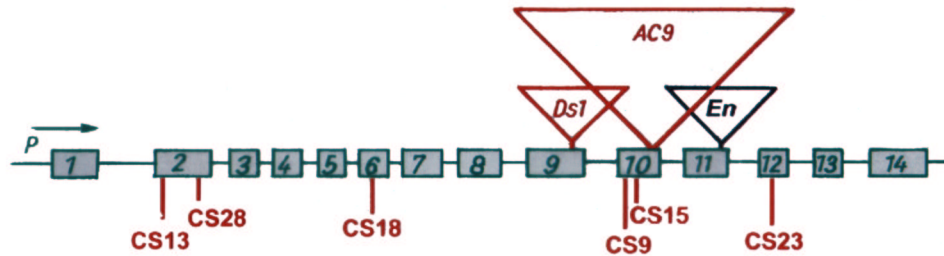


Abbildung 3: Mögliche Ergebnisse von Rekombinationsvorgängen zwischen homologen Transposonsequenzen an nicht allelen Positionen.

4.2.1 Klassifikation von Transposons und Transpositionsmechanismen

Transposons werden auf der Grundlage ihrer Sequenz und ihrer unterschiedlichen Transpositionsmechanismen klassifiziert [KUNZE et al. 1997, KIDWELL und LISCH 1997].

In der Klasse I werden Elemente zusammengefaßt, bei denen die Transposition über ein RNA-Zwischenprodukt verläuft. Dieses Zwischenprodukt wird durch Transkription eines im Genom vorliegenden Elements gebildet. Die Integration erfolgt durch eine reverse Transkription des RNA-Intermediats. Dieser Vorgang ist der Integration von Retroviren sehr ähnlich, daher werden Klasse I Elemente auch als Retrotransposons bezeichnet. Das Element, welches zur Erzeugung des Intermediats transkribiert wird, bleibt erhalten, weshalb die Transposition der Klasse I Elemente als replikative Transposition bezeichnet wird. Die Klasse I Transposons können weiter in *long terminal repeat* Elemente (LTR-Elemente) und non-LTR-Elemente eingeteilt werden. Die LTR-Elemente gleichen nicht nur hinsichtlich ihres Transpositionsmechanismus, sondern auch in ihrer Struktur den Retroviren. Zumindest bei einigen LTR-Transposons werden die RNA-Intermediate in virusartige Partikel verpackt, die in manchen Fällen retrovirale Funktionalität besitzen. Die non-LTR-Elemente werden weiter in *long interspersed elements* (LINE) und *short interspersed elements* (SINE) eingeteilt.



alleles	exon	DNA sequence	additional aminoacids
wt	9	ATC ACC GGC	0
Ds1-S5		ATC ACG GCC ATC ACC GGC	3
Ds1-S9		ATC ACC ATC ACC GGC	2
wt	10	GGA GAT	0
Ac9-r1		GGA GAT GGA GAT	2
wt	11	GTT	0
En-r1		GTT ATT	1
En-r2		GTC GTT	1

Abbildung 4: Mutanten des WAXY Gens aus Mais (*Zea mays* ssp. *mays*), die durch *footprints* erzeugt wurden. Alle hier gezeigten Sequenzvarianten codieren für ein enzymatisch aktives Protein.

Die Transposition von Klasse II Elementen erfolgt nach einem *cut and paste* Mechanismus, bei dem die Transposon-DNA ausgeschnitten und an der Integrationsstelle eingefügt wird. Bei der Integration von Klasse II Elementen werden kurze Duplikationen der angrenzenden Sequenz an den Elementgrenzen erzeugt. Als Ursache dafür wird die Erzeugung eines versetzten Doppelstrangbruchs an der Integrationsstelle angesehen.

Bei den meisten Transposonfamilien kommen autonome und nichtautonome Elemente vor. In der Sequenz autonomer Elemente sind die für die Transposition erforderlichen Faktoren codiert. Dies ist bei den nicht autonomen Elementen nicht der Fall, so daß diese nur mobil sind, wenn gleichzeitig ein autonomes Element im Genom vorhanden ist.

4.2.2 Erzeugung genetischer Diversität durch transposonbedingte Rearrangements

Transposons verändern die DNA-Sequenzen nicht nur bei Transpositionsvorgängen, sondern sie bieten darüber hinaus Ansatzpunkte für unterschiedlichste Formen der Flexibilisierung der Konfiguration des Genoms:

- In der Regel sind mehrere oder sogar überaus zahlreiche Kopien eines Elements in einem Genom zu finden. Dies gilt insbesondere für Retrotransposons, die einen erheblichen Teil der repetitiven DNA in vielen Genomen ausmachen. Kopien eines Transposons an verschiedenen Loci im Genom ermöglichen unsymmetrische homologe Rekombinationen. Durch diese können Teile von Chromosomen dupliziert werden oder verlorengehen (s. Abb. 3), es kann, je nach der relativen Orientierung der Transposons zueinander, zu Translokationen oder Inversionen kommen.
- Durch die von Retrotransposons codierte reverse Transkriptase können auch von anderen RNAs DNA-Kopien synthetisiert werden, die als intronfreie Pseudogene ins Genom integriert werden können. Diese wiederum können zum Ausgangspunkt der Evolution neuer Funktionalität werden.
- Transposons können Hotspots für induzierbare chromosomale Rearrangements sein. Beispielsweise treten bei Mais unter Streßbedingungen Chromosomenbrüche bei *double Ds* Elementen auf. Dies sind nichtautonome Elemente, die zur *Ac/Ds* Familie der Klasse II Elemente gehören.

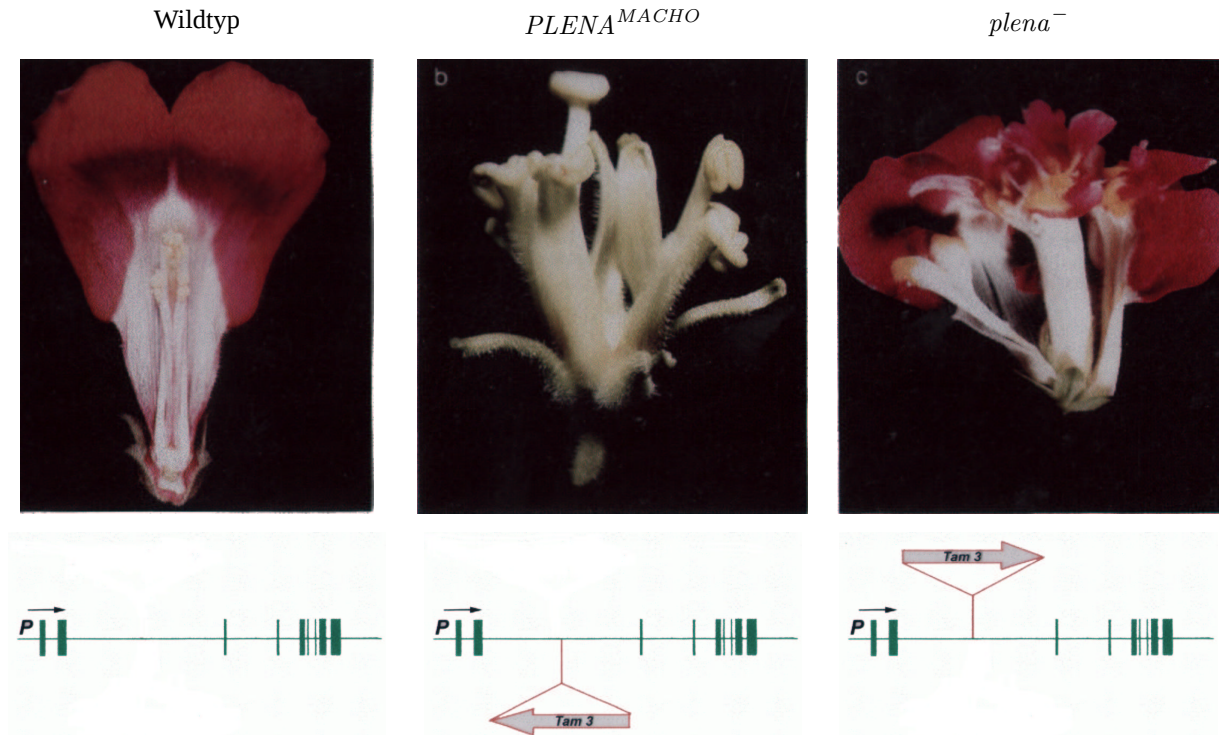


Abbildung 5: Oben: Wildtyp und komplementäre mutante Phänotypen der Blüte des Löwenmäulchens (*Antirrhinum majus*). Unten: Schematische Darstellung der Exon-Intron Struktur des *PLENA* Locus sowie der Position und Orientierung des integrierten *Tam 3* Elements bei den mutanten Allelen. Beide Mutationen sind durch eine Insertion des Transposons *Tam 3* in dasselbe Intron des *PLENA* Locus verursacht.

- Bei pflanzlichen Klasse II Transposons können bei der Excision verschiedene Varianten der ursprünglichen Sequenz zurückbleiben. Dies wird darauf zurückgeführt, daß die bei der Excision entstehenden Enden der DNA-Stränge für Exonucleasen und DNA-Polymerasen zugänglich sind. Das Exonucleasemodell der Excision [NEVERS et al. 1986] ist in Abb. 2 dargestellt. Die nach der Excision zurückbleibenden Veränderungen der Sequenz werden als *footprint* bezeichnet (s. Abb. 4).

4.2.3 Transposonbasierte Entstehung phänotypischer Diversität

Transpositionen und andere transposonbedingte Prozesse können unterschiedlichste Effekte auf die verschiedenen Ebenen der Realisierung biologischer Strukturen auf der Grundlage genetischer Information, also auf vielen phänotypischen Ebenen, haben. Oft spielen dabei besondere strukturelle Merkmale und andere Eigenschaften der beteiligten Transposons eine wesentliche Rolle.

4.2.3.1 Ebene der Aminosäuresequenzen Durch die Integration eines Transposons in den codierenden Bereich eines Gens wird in den meisten Fällen die Aminosäuresequenz des entsprechenden Genprodukts verändert. Ein häufiger Effekt ist die Expression unvollständigen und nicht funktionalen Proteins, verursacht durch ein Transkriptionsstopp-Signal auf dem Transposon oder durch eine Verschiebung des Translationsrasters.

Die *footprints*, die bei der Excision von pflanzlichen Klasse II Transposons zurückbleiben, können ebenfalls zu Veränderungen der Aminosäuresequenz führen. Wenn die Länge des *footprints* ein ganzzahliges Vielfaches von drei ist, wird das Leseraster nicht verschoben, sondern es werden nur eine oder wenige Aminosäuren in das Genprodukt eingefügt. Für das *WAXY*-Gen in Mais wurden etliche Mutanten dieser Art gefunden (Abb. 4). Die biologische Funktion so veränderter Proteine kann ganz oder teilweise erhalten bleiben, so daß der Phänotyp auf physiologischer und morphologischer Ebene als Reversion zum Wildtyp erscheint, trotz der Veränderung auf der Aminosäuresequenzebene.

Verschiedene Transposons enthalten Donor- und Akzeptorstellen für das RNA *splicing*. Durch die Integration solcher Transposons können Blöcke von Aminosäuren in ein Protein eingefügt oder aus ihm entfernt werden.

Es sind jedoch auch Integrationen in Introns möglich, bei denen die von dem betroffenen Gen codierte Aminosäuresequenz nicht verändert wird, weil mit dem Intron das gesamte Transposon beim *splicing* aus dem Primärtranskript herausgeschnitten wird. Derart integrierte Transposons bieten Ansatzpunkte für homologe Rekombinationen mit Transposonkopien an anderen Loci des Genoms. Durch ein solches Folgeereignis können codierende Bereiche unterschiedlicher Gene fusioniert werden, und somit können hybride Proteine evolvieren.

4.2.3.2 Andere phänotypische Ebenen Die vorangehend beschriebenen transposonbedingten Veränderungen von Aminosäuresequenzen können ihrerseits eine Vielzahl von teils hochkomplexen Effekten auf weiteren phänotypischen Ebenen hervorrufen. Beispielsweise spielen LTR-Retrotransposons bei der Entstehung vieler Tumoren beim Menschen eine wichtige Rolle.

Eine relativ direkte Folge der Expression nicht funktionaler Formen von Proteinen, die an der Synthese von Farbpigmenten beteiligt sind, ist das Fehlen des entsprechenden Pigments. Beispiele hierfür wurden in Abschnitt 2.5 (Abb. 1) gezeigt. Der Ausfall biologischer Proteinfunktionen kann jedoch auch weiterreichende Folgen bis zur Letalität haben; Mutationen in homöotischen Genen können z.B. die Morphologie einer Pflanze oder ihrer Blüten fundamental verändern (vgl. Abschnitt 4.3.2.1).

Auch durch Änderungen des Expressionsmusters eines Gens können Transposons drastische phänotypische Effekte verursachen. Ein Beispiel hierfür ist die Integration des Klasse II Elements *Tam 3* in ein Intron des *PLENA* Gens (s. Abb. 5). Das Genprodukt von *PLENA* ist ein Transkriptionsfaktor mit einer MADS-Domäne, dessen Expression in Primordien von Blütenorganen die Entwicklung reproduktiver Organe (Stamina und Karpelle) bestimmt (Genauerer hierzu s. Abschnitt 4.3.2.1). In Abwesenheit dieses Faktors entwickeln sich Organe des Perianths, also Kelch- und Kronblätter.

Bei der Integration von *Tam 3* in der Transkriptionsrichtung entgegengesetzter Orientierung wird dagegen das gesamte Transposon „rückwärts“ transkribiert. Beim anschließenden *splicing* der mRNA wird es zusammen mit dem Intron entfernt, so daß die prozessierte mRNA durch die Integration des Transposons nicht verändert wird. Bei dieser Mutante hat die Integration von *Tam 3* eine für die Regulation wichtige Sequenz im Intron des *PLENA* Gens zerstört. Das Gen wird infolgedessen konstitutiv expremiert, d.h. seine Expression ist nicht mehr auf die beiden inneren Wirtel der Blüte beschränkt. Die Integration von *Tam 3* in dieser Orientierung ergibt somit eine dominante Mutante, in der in allen Wirteln der Blüten reproduktive Strukturen gebildet werden (Abb. 5b).

Ist das Transposon in der Transkriptionsrichtung des Gens integriert, resultiert dies im vorzeitigen Abbruch der Transkription. Es kann damit keine vollständige mRNA und somit auch kein funktionales Genprodukt entstehen. Diese durch *Tam 3* verursachte Mutation ist rezessiv und führt phänotypisch zu Blüten, in denen Wirtel mit reproduktiven Strukturen völlig fehlen. Besonders auffällig ist dabei die Bildung von zwei Wirteln von Kronblättern (Abb. 5c).

4.2.4 Die Bedeutung von Transposons für Evolution und Biodiversität

Transposons werden bisweilen als genetische Parasiten bezeichnet [KIDWELL und LISCH 1997, BOEKE und DEVINE 1998]. Anders als bei Prokaryoten sind sie aber bei Eukaryoten allenfalls in Ausnahmefällen zu horizontalen Wirtswechseln in der Lage, ihre Vermehrung findet im Wesentlichen durch vertikale Transmission statt [BOEKE und DEVINE 1998]. Daher kommt es im Lauf der Koevolution von Transposon und Wirtsgenom in besonderem Umfang zu Adaptationen, welche die schädlichen Auswirkungen der Transposonaktivität für den Wirtsorganismus minimieren. Ergebnisse solcher Koadaptationen sind unter anderem eine Neigung zur Integration in nichtcodierenden Regionen des Genoms, die so ausgeprägt sein kann, daß große Ansammlungen ineinander verschachtelter Transposons in nichtcodierenden Regionen entstehen.

Transposons und andere repetitive Elemente in nichtcodierenden Regionen teilen ein Genom in Segmente ein, die Module biologischer Funktionalität codieren. Bei der aleatorischen Rekombination dieser Module ist die Rate der Erzeugung letaler oder ernsthaft dysfunktionaler Mutanten erheblich niedriger als bei ungerichteten Punktmutationen [CAPORAL 1999]. Auf diese Weise spielen Transposons eine wesentliche Rolle bei der Gestaltung genomischer Reorganisation im Laufe der Evolution.

Transposons ermöglichen damit die direkte Entstehung neuer Genomstrukturen, die auf dem Wege von Punktmutationen nur über zahlreiche durch Selektionsfaktoren benachteiligter Zwischenstufen erreicht werden können und gestatten dadurch eine strukturierte evolutionäre Exploration des Raums der Möglichkeiten genomischer Kon-

figurationen [CAPOREAL 1999]. So kann durch ein einziges Excisionsereignis eines pflanzlichen Klasse II Transposons ein *footprint* mit einer Länge, die ein ganzzahliges Vielfaches von drei ist, erzeugt werden. Somit kann mit einem Schritt der Code für ein längeres Protein entstehen, ohne daß Zwischenschritte mit Leserasterverschiebungen durchlaufen werden müssen (s. Abb. 4, welche mit hoher Wahrscheinlichkeit einer negativen Selektion ausgesetzt wären. Weiterhin wird die Evolution hybrider Proteine durch homologe Rekombination zwischen nichtallelen, repetitiven Sequenzen stark begünstigt. Auch die Integration von Transposons mit Donor- oder Akzeptorstellen für das *splicing* unterstützt die modulare Evolution von Proteinen.

Ein besonders illustratives Beispiel für die Bedeutung von Modulen biologischer Funktionalität für die Evolution sind bakterielle Plasmide, die multiple Antibiotikaresistenzen codieren. Die genetische Flexibilität, die der Geschwindigkeit der Evolution solcher Plasmide zugrundeliegt, wird durch IS-Elemente vermittelt [HU et al. 1975].

Alle Transposons benötigen für ihre Aktivierung neben den Faktoren, die auf den autonomen Elementen codiert sind, weitere vom Wirtsorganismus bereitgestellte Faktoren. Somit kann der Wirt die Aktivität seiner Transposons steuern (z.B. durch Methylierung [BANKS und FEDOROFF 1989]). In vielen Fällen wird bei günstigen Lebensbedingungen für den Wirtsorganismus eine geringe Transpositionsrate beobachtet, während verschiedene Streßfaktoren einen Anstieg der Transposonaktivität induzieren [NEVERS et al. 1986].

Zusammenfassend können Transposons somit als Systeme beschrieben werden, die den Umfang unspezifischer Diversität, die zur Evolution neuer biologischer Formen erforderlich ist, verringern und somit die Evolution von Biodiversität – im Sinne biologisch strukturierter Diversität – beschleunigen.

4.3 Biodiversität auf genetischer und morphologischer Ebene am Beispiel von Pflanzen

4.3.1 Drastische Unterschiede in der Diversität auf molekulargenetischer und morphologischer Ebene: das Fallbeispiel Teosinte-Mais

Ein besonders instruktives Beispiel für die Unterschiedlichkeit von Diversität auf molekularer und phänotypischer Ebene stellt die Gattung *Zea* (Mais und Teosintes) dar. Sie umfaßt mehrere Spezies, *Zea perennis*, *Zea diploperennis*, *Zea luxurians*, sowie *Zea mays*. *Zea mays* wird in die Subspezies *mexicana*, *parviglumis*, *huehuetenangensis* und *mays* eingeteilt, die mit Ausnahme von *Zea mays* ssp. *mays*, dem Kulturmais, alle als „Teosintes“ bezeichnet werden [DOEBLEY 1990b]. Es handelt sich hierbei um ursprünglich mittelamerikanische (vor allem mexikanische) Süßgräser. Studien auf der Grundlage molekularer Marker haben ergeben, daß die perennierenden Spezies einerseits (*Zea perennis* und *diploperennis*) und die verschiedenen *Zea mays*-Subspezies andererseits am wenigsten miteinander verwandt sind [DOEBLEY 1990a, BUCKLER und HOLTSFORD 1996]. Die Linien, die zu den entsprechenden Spezies geführt haben, haben sich wohl am frühesten in der Evolution voneinander getrennt. Auf molekularer Ebene repräsentieren die genannten Spezies damit das maximale Ausmaß an Divergenz innerhalb der Gattung *Zea*. Auf morphologischer Ebene zeigt sich aber ein ganz anderes Bild. Hier sehen alle Taxa (Teosintes) nahezu gleich aus, bis auf eine spezielle Subspezies, *Zea mays* ssp. *mays* (unser Kulturmais), die sich insbesondere im Gesamthabitus und der weiblichen Infloreszenz von den Teosintes stark unterscheidet (s. Abb. 7). Sie ist sehr nahe mit den anderen *Zea mays*-Subspezies verwandt. Eine extrem nahe Verwandtschaft besteht insbesondere mit *Zea mays* ssp. *parviglumis*, welches als unmittelbare Schwestergruppe des Kulturmais angesehen wird. Eine bestimmte Rasse dieser Subspezies, die im Balsas-Tal Mexicos vorkommt („Balsas-Teosinte“), gilt sogar als unmittelbarer „Ahne“ des Kulturmais [DOEBLEY 1990a].

Schauen wir uns die Unterschiede zwischen den Teosintes (einschließlich *Zea mays* ssp. *parviglumis*) und dem Kulturmais (*Zea mays* ssp. *mays*) etwas genauer an.

Morphologisch unterscheiden sich Teosinte und Mais in wenigstens fünf Merkmalskomplexen, die einerseits das Verzweigungsmuster und den Gesamthabitus der Pflanze, andererseits die weibliche Infloreszenz betreffen.

- i) Verzweigungsmuster: Teosinte ist eine Kurztagspflanze mit lateralem Verzweigungsmuster, wohingegen Mais eine Langtagspflanze mit einer ausgeprägten Hauptachse und lateral stark verkürzten Seitentrieben ist (Abb. 6). Alle terminalen Infloreszenzen der primären Seitenzweige von Teosinte sind männlich, wohingegen beim Mais lediglich die terminale Infloreszenz der Hauptachse männlich, die terminalen Blütenstände der gestauchten Seitentriebe jedoch weiblich sind.

Das stark ausgeprägte Verzweigungsmuster von Teosinte scheint in Mais unterdrückt zu sein. Es existiert jedoch eine Mais-Mutante, *tb1* („*teosinte branched 1*“), die wiederum ein Teosinte-ähnliches Verzweigungsmuster aufweist. *tb1*-Pflanzen besitzen terminale männliche Infloreszenzen an den Seitenzweigen, wie in Teosinte auch.

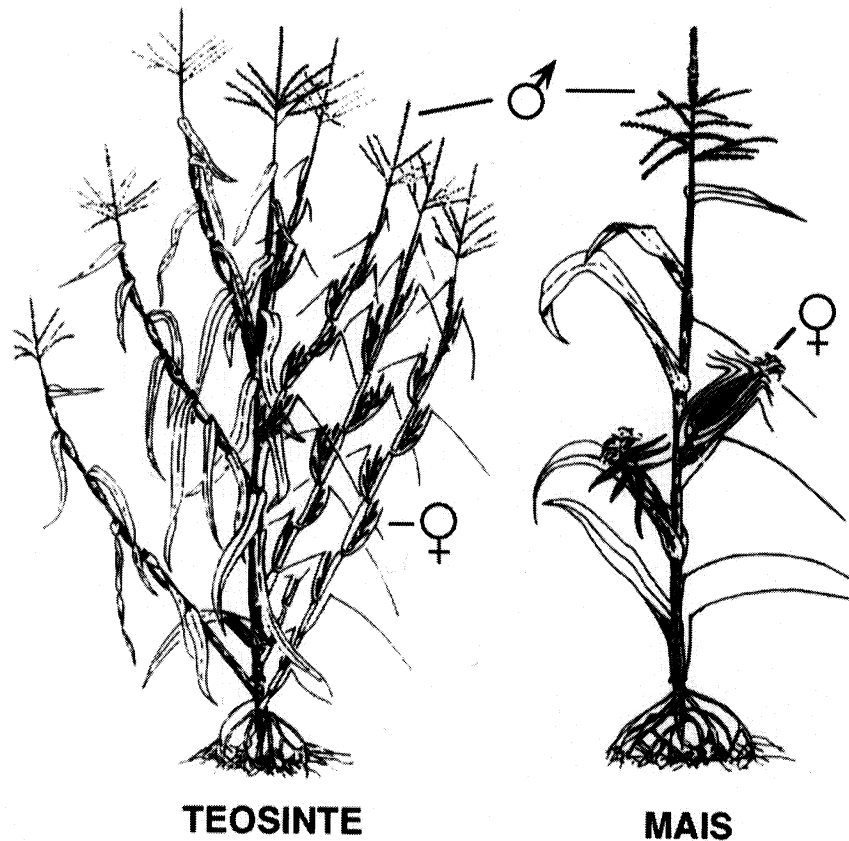


Abbildung 6: Habitus von Teosinte und Mais im Vergleich. Teosinte besitzt lange, primäre laterale Verzweigungen, welche von männlichen Infloreszenzen abgeschlossen werden. In Mais sind die primären lateralen Verzweigungen stark gestaucht; sie werden überdies von weiblichen Infloreszenzen abgeschlossen. Die weiblichen Infloreszenzen bei Teosinte entstehen aus sekundären lateralen Verzweigungen. Zur Vereinfachung wird im Text der Wachstumstyp von Teosinte als verzweigt, der von Mais als unverzweigt bezeichnet (nach: [DOEBLEY 1992], verändert).

- ii) Distichie vs. Polystichie: In Teosinte sind die Cupulen in zwei, gegenüberliegenden Reihen angeordnet, in Mais dagegen in mindestens vieren (Abb. 7 C u. D);
- iii) Einzelne vs. gepaarte Ährchen: In Teosinte stehen die Ährchen einzeln, während sie in Mais gepaart vorkommen (Abb. 7 C u. D).
- iv) brüchige vs. feste Ähre: im Reifezustand ist die Ähre von Teosinte durch den Besitz von Abscissionsschichten sehr fragil und erleichtert somit die Samenverbreitung, während die Maiskörner am Kolben verbleiben und somit zu ihrer Verbreitung auf den Menschen angewiesen sind (Abb. 7 A, E u. F).
- v) Verhärtete vs. weiche äußere Hüllspelzen: Teosinte besitzt langgestreckte, verhärtete, aufwärts-gerichtete äußere Hüllspelzen, während die entsprechenden Organe aus Mais verkürzt, weich und waagerecht orientiert sind (Abb. 7 E und F).

Aufgrund der großen morphologischen Unterschiede zwischen den Teosintes und Mais wurden beide Formen ursprünglich unterschiedlichen Gattungen zugeordnet (*Euchlaena* bzw. *Zea*) (für einen Überblick, siehe [SAEDLER und THEISSEN 1994])

Die Ursache des großen morphologischen Unterschiedes zwischen den Teosintes und Mais konnte in den letzten Jahren weitgehend aufgeklärt werden.

[BEADLE 1980] hat bereits durch Segregationsstudien einer Kreuzung zwischen Teosinte und einer primitiven Maissorte (Chapalote) festgestellt, daß die Merkmalsunterschiede zwischen Teosinte und Mais an nur fünf

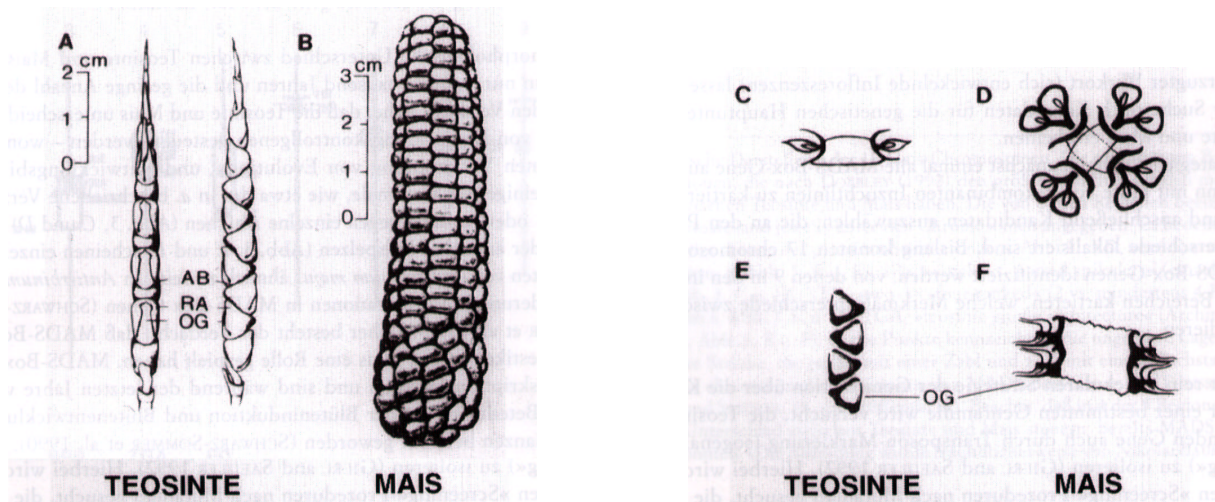


Abbildung 7: Neben dem Habitus (s. Abb. 6) unterscheiden sich Mais und Teosinte am meisten in der Architektur der weiblichen Infloreszenz. A: weibliche Infloreszenz von Teosinte; AB, Abscissionszone, welche zur Fruchtreife die Brüchigkeit der Ähre bewirkt; OG, äußere Hüllspelze (*outer glume*); RA, Internodium der Rachis. B: weibliche Infloreszenz von Mais („Maiskolben“). C, D: Schematische Querschnitte. C: Teosinte zeigt zwei Reihen Cupulen (Distichie) mit je einer Karyopse pro Cupule (einzelne Ährchen). D: Mais besitzt (mindestens) vier Reihen Cupulen (Polystichie) mit je zwei Karyopsen pro Cupule (gepaarte Ährchen). E, F: Längsschnitte. E: Vier Cupulen von Teosinte; man erkennt, daß die äußere Hüllspelze aufwärts, d.h. parallel zur Infloreszenzachse orientiert ist. F: Mais, bei dem die äußeren Hüllspelzen senkrecht zur Infloreszenzachse orientiert sind. (nach: [DOEBLEY 1992], verändert).

genomischen Loci kartieren, d.h. genetisch unterscheiden sich Mais und Teosinte kaum. [DOEBLEY 1992] hat entsprechende Experimente mit gleichem Ergebnis wiederholt, konnte aber durch Verwendung von RFLP-Kartierung die chromosomalen Positionen der Loci genauer bestimmen. Gemäß diesen Ergebnissen ist es naheliegend anzunehmen, daß nur ca. 5 Gene die grundlegenden morphologischen Unterschiede zwischen Teosinte und Mais kontrollieren.

Eines dieser Gene konnte als der schon erwähnte *tb1*-Locus identifiziert und anschließend kloniert werden [DOEBLEY et al. 1997]. Es zeigte sich, daß *TB1* wahrscheinlich für einen Transkriptionsfaktor kodiert, welcher das Wachstum von axillären Organen (Seitenzweigen) reprimiert. Bemerkenswerterweise unterscheiden sich die *TB1*-Allele von Teosinte und Mais nur relativ wenig. Das Mais-Allel wird stärker exprimiert als das Teosinte-Allel, was auf Veränderungen in der Promotor-Region des Genes zurückgeht [DOEBLEY et al. 1997, WANG et al. 1999] (s. Abschnitt 2.1).

Wir können an diesem Beispiel sehen, daß einzelne Gene, insbesondere solche, die die Entwicklung beeinflussen („Entwicklungskontrollgene“), einen drastischen Einfluß auf den Phänotyp haben können.

Obwohl die Genome von Mais und seinem wilden Vorfahren (Balsas-Teosinte) auf molekularem Niveau sehr divers sind, gibt es nur sehr wenig spezifische Diversität, die Mais und Balsas-Teosinte unterscheidet, nämlich die kleinen, lokalen Veränderungen in den entscheidenden Schlüsselgenen. Daher können wir einen drastischen Unterschied in der Divergenz auf gesamtgenomischen Niveau (wo wir eine spezifische Divergenz, welche Mais von Teosinte unterscheidet, von nahezu Null vorfinden) und phänotypischem Niveau (wo Balsas-Teosinte und Kulturmals eine Diversität zeigen, welche die Gesamtdiversität innerhalb der Gattung *Zea* nur wenig unterschreitet) konstatieren.

Solche Fälle kann man quantitativ nur beschreiben, wenn man ein quantitatives Mass für Biodiversität hat, das sich auf die verschiedenen Ebenen der Diversität anwenden läßt.

4.3.2 Vom Genom zum Phänotyp

Wir wollen jetzt die am Beispiel Teosinte-Mais illustrierten Schwierigkeiten, von der Diversität der Genome auf die phänotypische Diversität zu schließen, allgemeiner erörtern.

Zunächst ist zu klären, wie denn genomische Diversität molekularbiologisch gemessen bzw. definiert wird. Es wäre ideal, die Divergenz von Genomen durch den Vergleich der kompletten DNA-Sequenzen zu bestimmen, da bei einem solchen Verfahren die verfügbare genetische Information vollständig in die Divergenzmessung eingehen

würde. Die Entwicklung eines solchen Verfahrens ist jedoch problematisch, weil Genome einer komplexen evolutionären Dynamik unterliegen. Zur Ermittlung der Divergenz kompletter Genomsequenzen müßten diese Dynamik rekonstruiert werden, was auf der Grundlage der Sequenzinformation allein nur in begrenztem Umfang möglich ist.

Ferner wurden bislang lediglich einige relativ kleine Genome durchsequenziert (von Viren, Bakterien, Hefe und *C. elegans*), so daß die Datengrundlage für ein solches Verfahren (noch) nicht ausreicht. Daher bedient man sich derzeit in der Regel bestimmter Marker wie z.B. RFLPs, RAPDs, ISTRs, AFLPs etc. Diese Marker unterscheiden nicht zwischen funktionell wichtigen und unwichtigen Bereichen des Erbmaterials, d.h. sie ermittelt die durchschnittliche Divergenz des Gesamterbmaterials.

Um dies zu verstehen, wollen wir das Prinzip von zwei dieser Verfahren kurz erläutern. Beim RFLP (Restriktions-Fragment-Längen-Polymorphismus) mißt man den Abstand von Erkennungsstellen für Restriktionsendonucleasen im Genom, welche eine definierte Erkennungssequenz flankieren. Dazu verdaut man die gesamte genomische DNA eines Organismus mit einer oder mehreren Restriktionsendonucleasen, trennt die Fragmente mittels Gelelektrophorese auf, überträgt sie auf eine Membran, und weist dann Fragmente mit einer bestimmten Sequenz durch Hybridisierung mit einer spezifischen Sonde nach. Durch Vergleich mit einem Längenmarker kann man nun die Länge derjenigen Restriktionsfragmente bestimmen, welche die durch Hybridisierung nachgewiesene Sequenz tragen. Oftmals zeigen verschiedene Individuen unterschiedliche, und demnach charakteristische, Fragmentlängen, d.h. einen Restriktionsfragmentlängenpolymorphismus.

Beim RAPD (Random Amplified Polymorphic DNA) mißt man den Abstand zwischen zwei beliebigen, entgegengerichteten Primerbindungsstellen mittels PCR (*polymerase chain reaction*, Polymerasekettenreaktion) und geleelektrophoretischer Auftrennung der Produkte.

Inzwischen wurden intelligente Kombinationen der genannten Verfahren entwickelt, wie z.B. AFLP (*amplified restriction fragment length polymorphism*), welche die gleichzeitige Analyse sehr vieler Polymorphismen erlauben. Über die verschiedenen Verfahren liegen zahlreiche Publikationen vor (z.B. [CRAWFORD 1990, YU und PAULS 1992, VOS et al. 1995]). Diese Verfahren wurden bereits zur Bestimmung der genetischen Diversität bei diversen Pflanzen, u.a. Mais [LIVINI et al. 1992, DUBREUIL und CHARCOSSET 1998, SMITH et al. 1992, SMITH und SMITH 1992], Kakao [WHITKUS et al. 1998], Kokosnuß [LEBRUN et al. 1998] sowie Reis, [ZHU et al. 1998] eingesetzt.

Dieses setzt sich nun aus ganz unterschiedlichen Komponenten zusammen. Grob vereinfachend können wir Gene von nicht-funktionaler DNA unterscheiden. Gene kodieren für Proteine (Enzyme, Transkriptionsfaktoren, Signaltransduktoren etc.) oder funktionale RNA. Läßt man Phänomene wie Polyploidisierung etc. außer acht, so erscheint es wahrscheinlich, daß alle Blütenpflanzen eine etwa gleich große Anzahl an Genen besitzen, wobei 20.000 verschiedene Gene eine minimale Abschätzung ist. Diese Gene tragen in sehr unterschiedlichem Maße zur morphologischen und physiologischen Diversität bei. Gene, die für den pflanzlichen Grundstoffwechsel wichtig sind (z.B. Glykolyse, Citratzyklus, Photosynthese), spielen in den meisten Blütenpflanzen eine nahezu identische funktionale Rolle und tragen damit zur Diversität auf morphologischer oder physiologischer Ebene kaum etwas bei. Gene, die für den Sekundärstoffwechsel wichtig sind oder die Entwicklung steuern (wie das oben erwähnte Gen *TBI*) haben hingegen einen erheblichen Einfluß auf die strukturelle oder funktionale Diversität der Pflanzen. Manche Entwicklungskontrollgene haben einen derart bestimmenden Einfluß auf den morphologischen Phänotyp der Pflanze, daß man sie metaphorisch manchmal als „molekulare Architekten der Pflanzenbaupläne“ bezeichnet [THEISSEN und SAEDLER 1998]. Einige dieser Gene wollen wir uns exemplarisch etwas genauer ansehen.

4.3.2.1 MADS-Box-Gene als „molekulare Architekten der Blütenbaupläne“ Als besonders wichtige Schlüssel zum Verständnis der Entwicklungskontrolle pflanzlicher und tierischer Baupläne haben sich die sogenannten „homöotischen Gene“ erwiesen. Als solche bezeichnet man Gene, die bei einer Fehlfunktion zur Ausbildung von Organen an falschen Positionen des Tier- oder Pflanzenkörpers führen. Im Jahr 1990 wurde berichtet, daß mit *DEFICIENS* (*DEF*) aus dem Löwenmäulchen (*Antirrhinum majus*) und *AGAMOUS* (*AG*) aus dem Ackerschmalwand (*Arabidopsis thaliana*) die ersten homöotischen Gene aus Pflanzen isoliert worden waren [SOMMER et al. 1990, YANOFSKY et al. 1990]. Mutationen in diesen Genen führen zu homöotischen Veränderungen der Blütenstruktur. Die Blüten der höheren Angiospermen sind insofern segmental aufgebaut, als daß mehrere Organkreise (Wirtel) aufeinanderfolgen. Bei *Antirrhinum* und *Arabidopsis* liegen vier Wirtel vor, in denen (von Organkreis 1-4) Kelchblätter (Sepalen), Blütenkronblätter (Petalen), Staubblätter (Stamina) bzw. Fruchtblätter (Karpelle) sitzen. Bei einem Funktionsausfall von *DEF* werden im 2. Blütenkreis Kelchblätter statt Kronblätter und im 3. Blütenkreis Fruchtblätter statt Staubblätter ausgebildet. In der *AG*-Ausfallmutante erscheinen im 3. Blüten-

kreis Kronblätter anstelle von Staubblättern, und statt Fruchtblättern entwickeln sich im 4. Blütenkreis Kelchblätter bzw. iterativ neue Blüten. Derartige „gefüllte Blüten“ (vgl. hierzu auch Abschnitt 4.2.3, Abb. 5c) stellen bei vielen Zierpflanzen einen geschätzten Phänotyp dar.

Überraschenderweise zeigten die abgeleiteten Proteinprodukte von *DEFICIENS* und *AGAMOUS* in einer ca. 60 Aminosäurereste langen Region große Ähnlichkeit zu zwei gut charakterisierten Transkriptionsfaktoren aus Hefe (*MCM1* und *ARG80*) und dem „Serum response factor“ (SRF) der Säugetiere. Die Isolierung einiger weiterer homöotischer Gene aus Pflanzen bis zum Jahr 1992 erbrachte dann den ebenso bemerkenswerten Befund, daß von wenigen Ausnahmen abgesehen, diese auch alle über die konservierte Region verfügen [DAVIES und SCHWARZ-SOMMER 1994]. Nach den vier „Gründungsgenen“ *MCM1*, *AGAMOUS*, *DEFICIENS* und *SRF* wurde die in allen Genen hochkonserviert vorliegende Region MADS-Box genannt [SCHWARZ-SOMMER et al. 1990]. Ihre Ähnlichkeit zu den MADS-Box-Genen aus Tieren und Hefe und die Fähigkeit ihrer Translationsprodukte, an DNA zu binden, deuten darauf hin, daß auch die pflanzlichen MADS-Box-Gene für Transkriptionsfaktoren kodieren (für einen aktuellen Überblick über pflanzliche MADS-Box-Gene, siehe Theißen und Saedler, 1995; Theißen et al., 1996, 1999; Riechmann und Meyerowitz, 1997).

Die Funktion homöotischer Gene ist an ihrem mutanten Phänotyp erkennbar. Dieser kann sehr verschieden sein, abhängig davon, ob ein Gen funktionell ausfällt oder falsch (ektopisch) exprimiert wird. Beispielsweise führt der Ausfall des Genes *PLENA*, dem Orthologen des *Arabidopsis*-Gens *AG*, zu gefüllten Blüten wie bei der *AG*-Mutante, d.h. es werden ausschließlich sterile Blütenhüllblätter gebildet (s. Abb. 5). Wird hingegen dasselbe Gen nicht nur im 3. und 4. Blütenkreis (wie im Wildtyp), sondern auch im 1. und 2. Blütenkreis ausgeprägt, erhält man den komplementären Phänotyp (*MACHO/OVULATA*) einer ausschließlich aus Reproduktionsorganen bestehenden Blüte, bei der Fruchtblätter im 1. und 4. und Staubblätter im 2. und 3. Blütenkreis gebildet werden (vgl. Abschnitt 4.2.3, Abb. 5b) [BRADLEY et al. 1993]. Aus diesen und ähnlichen Befunden kann man ableiten, daß die entsprechenden homöotischen Gene die Identität von Blütenorganen festlegen. Sie werden deshalb auch „Blütenorganidentitätsgene“ genannt und zeigen damit allgemeine funktionelle Analogien zu den homöotischen Selektorgenen vom *HOX*-Typ der Tiere.

Zu Beginn der 90er Jahre wurden auch die ersten Klonierungen von Mitgliedern einer weiteren Klasse von Pflanzengenen, der „Blütenmeristemidentitätsgene“ oder „Heterochroniegene“ berichtet. Da die Expression dieser Gene Meristemen eine florale Identität verleiht, kommt es bei deren Funktionsverlust zu einer teilweisen oder vollständigen Umwandlung von Blüten in Infloreszenzen. Weil Ort und Zeitpunkt von Entwicklungsvorgängen in Pflanzen stark miteinander korreliert sind, kann ein solcher Phänotyp sowohl als homöotische, als auch als heterochrone Veränderung aufgefaßt werden. Bei letzterer wird ein bestimmtes Merkmal zu einem falschen Zeitpunkt in der Individualentwicklung ausgeprägt. Beispielsweise entwickeln sich in *SQUAMOSA*-Mutanten von *Antirrhinum* sekundäre Infloreszenzen an einer Stelle (homöotischer Effekt) und zu einem Zeitpunkt (heterochroner Effekt), wo im Wildtyp Blüten gebildet werden [HUIJSER et al. 1992]. Auch viele der bislang isolierten Blütenmeristemidentitätsgene, wie z.B. *SQUAMOSA* (*SQUA*) aus *Antirrhinum majus* und *APETALA1* (*AP1*) aus *Arabidopsis thaliana*, erwiesen sich als MADS-Box-Gene [HUIJSER et al. 1992, MANDEL et al. 1992].

Seit der Isolierung der ersten homöotischen und heterochronen Gene aus Pflanzen wurden weitere große Fortschritte im Verständnis der molekulargenetischen Grundlage der Blütenentwicklung gemacht. Dadurch ist die Blütenentwicklung einer der bestverstandenen morphogenetischen Prozesse bei Pflanzen geworden.

Insbesondere durch Untersuchungen an *Arabidopsis* ist mittlerweile klar geworden, daß die homöotischen und heterochronen Gene in ein komplexes Gen-Netzwerk integriert sind, welches einen wesentlichen Beitrag zur Blütenentwicklung leistet. Die Bestandteile dieses Netzwerkes kodieren überwiegend für Transkriptionsfaktoren. Dieses Gen-Netzwerk hat unverkennbare Ähnlichkeit zu analogen Gen-Netzwerken, die die Entwicklung der Körperbaupläne der vielzelligen Tiere steuern [THEISSEN und SAEDLER 1995]. Beiden Typen von Gen-Netzwerken gemeinsam ist z.B. ein hierarchischer Aufbau aus mehreren Ebenen, bei denen ein Teil der höherpostierten Gene Positionsinformation etabliert, die dann von einer tieferen Ebene homöotischer Selektorgene in Form von definierten Kompartimenten manifestiert wird. Die Selektorgene wiederum aktivieren „Realisatorgene“, deren Genprodukte letztlich für den Phänotyp (z.B. „Organidentität“) der verschiedenen Kompartimente verantwortlich sind (für einen konzeptionellen Überblick, siehe [TAUTZ 1996]; das Gen-Netzwerk der Blütenentwicklung wird von [THEISSEN und SAEDLER 1995] beschrieben). Im Fall der Blütenentwicklung bei *Arabidopsis* bewirken nach ihrem mutanten Phänotyp benannte *late and early flowering genes* („Gene für frühe und späte Blüte“, auch *flowering time genes* („Blühzeitpunktgene“) genannt) das Umschalten vom vegetativen zum generativen Wachstum, teilweise in Abhängigkeit von Umweltfaktoren. Die Expression der schon erwähnten Blütenmeristemidentitätsgene verleiht den an den Flanken der Infloreszenzachse entstehenden Blütenmeristemen ihre florale Identität. Durch die

Aktivität sogenannter *cadastral genes* („Katastergene“) wird Positionsinformation für vier verschiedene Blütenkreise etabliert. Die kombinatorische Interaktion von drei verschiedenen „homöotischen Genfunktionen“ verleiht dann den verschiedenen Blütenkreisen ihre spezifische Identität. Nach dem sogenannten „ABC-Modell“ wird im ersten und zweiten Blütenkreis die A-Funktion exprimiert, im zweiten und dritten Blütenkreis die B-Funktion, und im dritten und vierten Blütenkreis die C-Funktion. Dabei determiniert die Expression der A-Funktion im ersten Blütenkreis die Entwicklung von Kelchblättern (Sepalen), die Co-Expression der A- und B-Funktion im zweiten Blütenkreis bewirkt die Entwicklung von Kronblättern (Petalen), die Co-Expression der B- und C-Funktion im dritten Blütenkreis determiniert die Entwicklung von Staubblättern (Stamina), und die Expression der C-Funktion im vierten Blütenkreis bewirkt, daß sich hier Fruchtblätter (Karpelle) entwickeln (einen Überblick über das ABC-Modell bieten [WEIGEL und MEYEROWITZ 1994, MEYEROWITZ 1998, THEISSEN und SAEDLER 1995]). Das ABC-Modell stützt sich insbesondere auf die Analyse von homöotischen Mutanten, von denen weiter oben bereits zwei verschiedene Typen exemplarisch beschrieben wurden. Fällt beispielsweise die B-Funktion aus (wie z.B. im Falle der *DEF*-Mutante), erscheinen Sepalen im ersten und zweiten Blütenkreis, und Karpelle im dritten, entsprechend der alleinigen Expression der A-Funktion in den ersten beiden und der C-Funktion in den letzten beiden Blütenkreisen. Im strengeren Sinne gültig ist das ABC-Modell aber nur für *Arabidopsis*, schon bei *Antirrhinum* und *Petunia* treten Abweichungen auf [THEISSEN et al. 1996, THEISSEN et al. 2000]. Seine Anwendbarkeit auf entfernter verwandte Taxa wird zur Zeit untersucht [THEISSEN et al. 2000].

Die homöotischen Genfunktionen A, B und C werden von den bereits besprochenen „Blütenorganidentitätsgenen“ bereitgestellt, die den „Selektorgenen“ der Tiere entsprechen. *Arabidopsis*-Mutanten in allen drei homöotischen Funktionen sind bekannt, und die entsprechenden Gene konnten kloniert werden. Die A-Funktion wird von den Genen *APETALA1* (*AP1*) und *APETALA2* (*AP2*) bereitgestellt, die B-Funktion von *APETALA3* (*AP3*) und *PISTILLATA* (*PI*), und die C-Funktion von dem bereits erwähnten Gen *AGAMOUS* (*AG*). Mit Ausnahme von *AP2* sind diese Gene Mitglieder der MADS-Box-Genfamilie. In *Antirrhinum* wird die B-Funktion von den beiden Genen *DEFICIENS* (*DEF*) und *GLOBOSA* (*GLO*), und die C-Funktion von *PLENA* (*PLE*) bereitgestellt, wobei es sich in allen drei Fällen um MADS-Box-Gene handelt [DAVIES und SCHWARZ-SOMMER 1994, WEIGEL und MEYEROWITZ 1994]. Ob *Antirrhinum* über eine A-Funktion verfügt, wurde kontrovers diskutiert (vgl. [DAVIES und SCHWARZ-SOMMER 1994] mit [WEIGEL und MEYEROWITZ 1994]). Eine Klärung der Kontroverse kann möglicherweise durch die Berücksichtigung des evolutionären Ursprungs der A-Funktion erfolgen [THEISSEN et al. 2000].

Ähnlich den Gen-Netzwerken, die die tierische Embryogenese steuern, setzen sich die analogen Netzwerke der Blütenentwicklung aus Mitgliedern weniger Genfamilien zusammen. In diesen spielen MADS-Box-Gene eine wichtige Rolle, und zwar auf mehreren Ebenen. Wie schon erwähnt stammt neben der überwiegenden Mehrzahl der Blütenorganidentitätsgene auch ein Teil der floralen Meristemidentitäts- und Katastergene aus der Familie der MADS-Box-Gene (s. [DAVIES und SCHWARZ-SOMMER 1994, WEIGEL und MEYEROWITZ 1994]; für einen Überblick, siehe [THEISSEN und SAEDLER 1995]). Bestimmte *AG*-ähnliche MADS-Box-Gene konnten in Petunie auch als *master control genes* („oberste Kontrollgene“) der Ovula-Entwicklung identifiziert werden [COLOMBO et al. 1995]. Die entsprechenden Gene sind in einer Erweiterung des ABC-Modells als „D-Funktionsgene“ bezeichnet worden. Darüberhinaus sind MADS-Box-Gene vermutlich auch an der Steuerung vieler anderer Entwicklungsprozesse dikotyler Pflanzen beteiligt, wie Embryogenese und vegetative Entwicklung (für einen Überblick, siehe [THEISSEN und SAEDLER 1995, THEISSEN et al. 1996, THEISSEN et al. 2000]). Neben MADS-Box-Genen haben unter anderem auch *APETALA2*-ähnliche Gene, *LEAFY/FLORICAULA*-ähnliche Gene und Homöobox-Gene wichtige Funktionen in der Blütenentwicklung (für einen Überblick, siehe [THEISSEN und SAEDLER 1995, THEISSEN und SAEDLER 1998, THEISSEN et al. 1996]).

4.4 Schlußfolgerungen

Wir haben gesehen, daß verschiedene Teile des Genoms einen ganz unterschiedlichen Einfluß auf den Phänotyp und damit die phänotypische Diversität haben können. Als Extreme haben wir einerseits Transposons und andererseits Entwicklungskontrollgene kennengelernt. Transposons und andere repetitive Elemente sind für die Entwicklung weitestgehend unbedeutend, für die evolutionäre Dynamik haben sie dagegen eine entscheidende Bedeutung, indem sie Genome modularisieren und dadurch die Evolution neuer biologischer Formen beschleunigen. An einem Beispiel (*Tam 3* im *PLENA*-Gen, Abb. 5) haben wir aber auch gesehen, daß Transposons die Funktion von Entwicklungskontrollgenen direkt und massiv beeinflussen können. Entwicklungskontrollgene (*TB1*, bestimmte MADS-Box-Gene) sind für die Morphogenese von zentraler Bedeutung, sie entfalten ihren Einfluß auf den mor-

phologischen Phänotyp über ein komplexe Netz von Wechselwirkungen.

Sowohl die Phylogenese als auch die Ontogenese der Lebensformen in ihrer heutigen morphologischen Diversität sind somit als Resultate der Dynamik genetischer Information anzusehen. Daher erscheint es verfehlt, sich bei der Definition und Untersuchung von Biodiversität auf die phänotypische Ebene zu konzentrieren und allenfalls die „für den Phänotyp relevanten Gene“ in die Betrachtung einzubeziehen. Vielmehr sollte man die Tatsache ausnutzen, daß das Messen von Diversität auf Sequenzebene bzw. der Ebene molekularer Marker mithilfe einfacher und wohldefinierter Verfahren, wie z.B. die in Abschnitt 4.3.2 kurz erläuterten Verfahren RFLP, RAPD und AFLP, möglich ist. Da den verglichenen Genomabschnitten bei diesem Vorgehen keine funktionale Bedeutung zugewiesen werden muß und auch sonst keine Annahmen oder Klassifikationen, die implizit anthropogene Einflüsse in die Analysen einbringen können, erforderlich sind, zeichnen sich diese Verfahren durch ein großes Maß an Objektivität aus. (Wir sehen hier eine interessante Parallele zur Informationstheorie: während Information auf syntaktischer Ebene leicht zu definieren und zu messen ist, tut man sich mit der quantitativen Messung von Information auf der semantischen Ebene sehr schwer.)

Weil wir uns beim Phänomen der Biodiversität aber oftmals gerade für die phänotypischen Aspekte der Diversität interessieren, und im Kontrast zwischen genotypischer und phänotypischer Diversität ein interessantes Phänomen ausgemacht haben, suchen wir nach einem Verfahren, Biodiversität auf allen Ebenen messen und dann vergleichen zu können.

5 Ein Ansatz zur quantitativen Charakterisierung von Biodiversität

5.1 Distanzen

Grundlagen jedes Diversitätskonzepts sind stets eine Menge mit mehreren Objekten sowie eine Möglichkeit, die Unterschiedlichkeit von Objekten in dieser Menge quantitativ zu bestimmen. Das formale Instrument zur quantitativen Beschreibung von Unterschiedlichkeit sind Distanzmaße.

Distanzmaße sind Verfahren, mit denen für Paare von Objekten skalare, nichtnegative Werte ermittelt werden. Voraussetzung dafür ist, daß die Objekte vergleichbare Eigenschaften oder Merkmale aufweisen, die unterschiedlich ausgeprägt sein können. Die Distanz zwischen zwei Objekten beschreibt quantitativ das Ausmaß der Unterschiedlichkeit hinsichtlich der betrachteten Eigenschaften.

Fast alle biologischen Systeme, von der Zelle bis zum Ökosystem, sind Organisationen aus vielen Untereinheiten, die vergleichbare Eigenschaften besitzen. Eine Charakterisierung von Systemen auf der Grundlage von Distanzmaßen ist somit auf allen Ebenen biologischer Organisation, die diese Eigenschaft aufweisen, möglich. Die mit solchen Maßen ermittelten Distanzen enthalten Information über die evolutionären Prozesse, die der Entstehung der untersuchten Einzelkomponenten des betrachteten biologischen Systems zugrundeliegen (vgl. Abschnitt 4.1). Distanzbasierte Verfahren sind somit zur quantitativen Charakterisierung von Biodiversität geeignet. Insbesondere bieten sie die Möglichkeit, Biodiversität auf verschiedenen Ebenen biologischer Organisation zu ermitteln und zu vergleichen.

Wir untersuchen und vergleichen in unserem Beitrag daher hauptsächlich distanzbasierte Biodiversitätsmaße. Um eine möglichst weitgehende Universalität dieses Ansatzes zu gewährleisten, betrachten wir schwerpunktmäßig Distanzmessungen zunächst auf der Grundlage genetischer Information, berücksichtigen aber auch Distanzen, die auf der Phänotyp-Ebene und auf anderen Ebenen biologischer Organisation ermittelt werden können.

5.2 Diversitätsmaße auf der Grundlage von Distanzen

Die Bedeutung von evolutionsbasierten Distanz- oder Divergenzmessungen für die quantitative Charakterisierung von Biodiversität wurde bereits in verschiedenen Arbeiten erwähnt. Die folgenden Abschnitte geben einen Überblick über einige der vorgeschlagenen Ansätze und Maße, wobei kein Anspruch auf Vollständigkeit erhoben wird. Die hier vorgestellten Biodiversitätsmaße wurden von den jeweiligen Autoren zur Anwendung auf der taxonomischen Ebene entwickelt, lassen sich aufgrund ihrer formalen Struktur aber ohne Schwierigkeiten auch auf anderen Ebenen biologischer Organisation einsetzen.

5.2.1 Die Biodiversitätsmaße von WILLIAMS, HUMPHRIES und VANE-WRIGHT

Das Prinzip der quantitativen Charakterisierung auf der Grundlage phylogenetischer Bäume und daraus abgeleiteter Distanz- oder Divergenzmaße wird erstmalig in [WILLIAMS et al. 1991] eingeführt. In diesem Artikel werden auch verschiedene Biodiversitätsmaße vorgeschlagen und anhand von Beispielen mit der Quantifizierung von Biodiversität durch die Artenzahl verglichen.

Grundlage aller in [WILLIAMS et al. 1991] vorgeschlagener Biodiversitätsmaße ist die Topologie des Stammbaums der in einem System (z.B. einem Habitat) lebenden Spezies. Diese Spezies werden durch die Endknoten eines bewurzelten, vollständig aufgelösten Stammbaums repräsentiert. Die Gesamtzahl der Spezies wird im Folgenden mit n bezeichnet, die einzelnen Spezies seien mit dem Index $i \in \{1, \dots, n\}$ durchnummeriert. Anhand der Stammbaumtopologie werden nun einige Basisgrößen definiert:

- I_i ist die Anzahl der auf dem Pfad vom Endknoten i zur Wurzel durchlaufenen Knoten.
- S_{jk} ist die Anzahl der Knoten, die sowohl auf dem Pfad vom Endknoten j zur Wurzel als auch auf dem Pfad vom Endknoten k zur Wurzel liegen (*shared nodes*).
- U_{jk} ist die Anzahl der Knoten, die auf dem Pfad vom Endknoten j , aber nicht auf dem Pfad vom Endknoten k zur Wurzel liegen (*unique nodes*).

Von diesen Basisgrößen werden nun das *root weight* (Gl. 1) und verschiedene Divergenzmaße abgeleitet. Die im Folgenden verwendeten römischen Indices sind aus [WILLIAMS et al. 1991, Seite 672] übernommen:

$$R(i) = \frac{1}{I_i} \quad (1)$$

$$d_{III}(j, k) = \frac{1}{S_{jk}} \quad (2)$$

$$d_{IV}(j, k) = U_{jk} + U_{kj} + 1 \quad (3)$$

$$d_V(j, k) = \frac{U_{jk} + U_{kj} + 1}{2S_{jk} + U_{jk} + U_{kj}} \quad (4)$$

$$d_{VI}(j, k) = \frac{U_{jk} + U_{kj}}{S_{jk}} + 1 \quad (5)$$

Bemerkenswert ist hierbei, daß d_{IV} nicht von der Wurzelposition abhängt, obwohl U_{jk} wurzelpositionsabhängig ist; d_{IV} mißt die Anzahl interner Knoten auf dem Pfad zwischen den Endknoten j und k .

Die Divergenzmaße werden nunmehr zur Berechnung von Biodiversitätsmaßen verwendet. Zur Verwendung des *root weight* als Biodiversitätsmaß werden die Werte für die einzelnen Spezies aufsummiert:

$$m_{II} = \sum_i R(i) \quad (6)$$

Vier Maße sind jeweils durch den mit der Speziesanzahl gewichteten Mittelwert der Divergenz definiert:

$$m_x = n \cdot \frac{2}{n(n-1)} \cdot \sum_{i < j} d_x(i, j), \quad x = III, IV, V, VI \quad (7)$$

Die taxonomische Dispersion (*taxonomic dispersion*) ist schließlich folgendermaßen definiert:

$$m_{VII} = n \cdot (\text{Mittelwert der } d_{IV}) \cdot (\text{Standardabweichung der } d_{IV}) \quad (8)$$

Alle hier genannten Maße werden nun dazu verwendet, den relativen Verlust an Biodiversität zu charakterisieren, wenn einige Spezies aussterben und somit nur eine Teilmenge der ursprünglichen Speziesmenge erhalten bleibt. Die erhaltene Biodiversität wird jeweils durch

$$\frac{m_x(\text{Teilmenge})}{m_x(\text{Gesamtmenge})} \cdot 100\%, \quad x = II, III, IV, V, VI, VII \quad (9)$$

quantifiziert. Ein relevantes Detail hierbei ist, daß die Autoren dabei für die Ermittlung der Divergenzwerte innerhalb der Teilmenge und in der Gesamtmenge denselben Stammbaum verwenden. Dies bedeutet, daß bei den Pfadlängenberechnungen auch Knoten, die im Stammbaum der Teilmenge keine Verzweigungen darstellen, in die Berechnung von m_x (Teilmenge) eingehen.

Zentrales Motiv bei der Entwicklung dieser Biodiversitätsmaße ist es, phylogenetische Information zu verwenden, um den Beitrag zur Biodiversität, der von Arten mit einem größeren evolutionären Abstand vom Rest des Artenspektrums kommt, stärker zu gewichten.

5.2.2 Formale Kriterien für Biodiversitätsmaße

SOLOW, POLASKY und BROADUS halten das Ausmaß der Unterschiedlichkeit, also die Distanz, einer Art zu den anderen Arten eines Ensembles ebenfalls für relevant für die Charakterisierung von Biodiversität [SOLOW et al. 1993]. Diese Autoren schlagen neben speziellen Biodiversitätsmaßen einige Eigenschaften vor, die ein distanzbasiertes Biodiversitätsmaß erfüllen sollte:

- Monotonität: Die Diversität einer Teilmenge (von Spezies) soll kleiner oder höchstens gleich der Diversität der Gesamtmenge der Spezies sein.
- Das Hinzufügen einer Spezies, welche zu (mindestens) einer der in der Speziesmenge bereits vorhandenen Spezies die Distanz 0 hat, soll die Diversität nicht erhöhen.
- Die Zunahme der Diversität einer Speziesmenge beim Hinzufügen einer Spezies soll mit der kleinsten Distanz, die die Spezies zu einer in der Menge bereits enthaltenen Art hat, in positiver Weise korreliert sein.

5.2.3 Die Baumlänge als Biodiversitätsmaß

NEE und MAY schlagen als Maß für die Biodiversität die Länge eines Stammbaums vor [NEE und MAY 1997]. Diese ist definiert als die Summe der Längen aller Kanten in einem Stammbaum. Die Autoren beziehen sich dabei auf ultrametrische Stammbäume, bei denen die Kantenlängen evolutionäre Zeitabstände repräsentieren. Das Konzept kann jedoch ohne weiteres auch bei additiven, nicht ultrametrischen Stammbäumen, wie man sie bei Phylogenierekonstruktionen auf der Grundlage molekulargenetischer Information erhält, verwendet werden. Die Baumlänge ist insofern als distanzbasiertes Biodiversitätsmaß zu betrachten, als die Kantenlängen Distanzen zwischen den Knoten des Stammbaums darstellen.

5.2.4 Stammbaumunabhängige distanzbasierte Biodiversitätsmessung

Der gewichtete Mittelwert der Divergenz (Gl. 7, Abschnitt 5.2.1) und die taxonomische Dispersion (Gl. 8) sind Biodiversitätsmaße, die auf der Grundlage jeder Matrix von Distanz- oder Divergenzwerten berechnet werden können. Es ist für die Berechnung dieser Maße irrelevant, ob diese Werte auf der Grundlage einer Stammbaumtopologie ermittelt wurden oder aus anderen Quellen stammen.

Bei der Stammbaumrekonstruktion auf der Grundlage molekularer Daten zählen distanzbasierte Verfahren zu den Standardmethoden (s. [FELSENSTEIN 1989, LI und GRAUR 1991, FELSENSTEIN 1995]). Bei diesen Verfahren wird aus den empirisch erhobenen Meßdaten (z.B. Sequenzen oder RFLP-Daten) zunächst eine Distanzmatrix berechnet, wobei die Distanzen Schätzwerte für die seit der Divergenz zweier Entitäten verstrichenen Zeit sein sollen. Diese Distanzmatrix dient als Eingabe für Verfahren zur Rekonstruktion eines Stammbaums.

Das Errechnen von Schätzwerten für die seit der Divergenz vergangenen Zeit ist auf der Grundlage molekularer Daten nur unter der Annahme neutraler Evolution möglich. Diese Annahme ist jedoch stets nur näherungsweise erfüllt, so daß die Schätzwerte systematische Verzerrungen aufweisen, die zu verschiedensten Problemen bei der Stammbaumrekonstruktion führen.

Daher bietet es sich an, die distanzbasierten Biodiversitätsmaße unmittelbar auf die auf der Basis molekularer Daten berechneten Distanzen anzuwenden, da auf diese Weise der problematische Stammbaumrekonstruktions-schritt vermieden wird. Aus diesem Grund wurde das gewichtete Divergenzmittel solcher Distanzmatrizen in die empirischen Untersuchungen (Abschnitt 6) einbezogen. Auch die Distanzverteilungskomplexität (s. Abschnitt 5.3) wurde auf der Grundlage solcher Distanzwerte gemessen.

5.2.5 Weitere Biodiversitätsmaße

In der Literatur wurde neben den vorangehend genannten Maßen noch eine Vielzahl weiterer distanzbasierter Biodiversitätsmaße vorgestellt. Beispiele hierfür sind die cladistische Dispersion und die genetische Diversität, s. [CROZIER und KUSMIERSKI 1994]. Diese Maße wurden nicht in die im folgenden Abschnitt vorgestellten Untersuchungen einbezogen, u.a. weil sie sich nur bedingt zum Einsatz mit molekularen Daten eignen.

5.3 Die Distanzverteilungskomplexität (DVK) als Biodiversitätsmaß

Unter Biodiversität kann man allgemein die Vielfalt beliebiger Mengen biologischer Objekte verstehen. In diesem Abschnitt soll darunter die Diversität von solchen biologischen Objekten verstanden werden, die gemeinsam evolviert sind.

Systeme mit Biodiversität in diesem Sinne unterscheiden sich von Systemen, in denen es keine oder kaum Diversität gibt (z.B. Monokulturen). Weiterhin sind sie aber auch von Systemen mit unstrukturierter Diversität zu unterscheiden, da biologische Diversität stets durch komplexe Selektionsprozesse strukturiert ist. Biodiversitätsmaße sollten daher sowohl für monokulturartige Systeme als auch für Systeme mit unstrukturierter Diversität niedrigere Werte als für Systeme mit biologisch strukturierter Diversität liefern.

Biologische Selektionsmechanismen erzeugen in vielen Fällen hierarchische Strukturen. So lassen sich z.B. Gene in Genfamilien und Subfamilien einteilen, und verschiedene Lebensformen können in Domänen, Reiche, Stämme usw. bis hin zu Spezies und Subspezies eingeteilt werden. Diese Strukturen können mithilfe von Distanzmaßen gut erfaßt werden. Ein Beispiel hierfür sind die distanzbasierten Methoden für die Phylogenierekonstruktion.

In der Komplexitätsforschung werden Maße, die sowohl hochgradig geordneten Systemen als auch hochgradig ungeordneten Systemen niedrige Komplexitätswerte zuweisen, als Komplexitätsmaße 2. Ordnung (oder auch als alternative Komplexitätsmaße [KURTHS et al. 1995]) bezeichnet. Die Distanzverteilungskomplexität (DVK) [KIM 1996b, KIM 1996a] ist ein distanzbasiertes Komplexitätsmaß zweiter Ordnung. Ursprünglich wurde sie zur quantitativen Charakterisierung der taxonomischen Komplexität bei Simulationen evolvierender Populationen entwickelt. Da taxonomische Komplexität mit biologisch strukturierter Vielfalt in sehr enger Korrelation stehen (man kann sogar fragen, ob es sich möglicherweise zwei Begriffe für dasselbe Phänomen handelt) ist die DVK als Biodiversitätsmaß geeignet, zumindest bietet sie sich als Ausgangspunkt für die Entwicklung eines Biodiversitätsmaßes an.

Die DVK erfüllt die in Abschnitt 5.2.2 genannten Monotonitätsforderung nicht. Es ist daher möglich, daß die DVK als alleiniges Maß zur quantitativen Charakterisierung von Biodiversität nicht geeignet ist. Sie kann aber in Kombination mit anderen Maßen dazu dienen, die biologische Struktur von Diversität zu charakterisieren.

5.3.1 Definition der Distanzverteilungskomplexität

Die DVK ist definiert als die Shannon-Entropie der Verteilung der paarweisen Distanzen der Individuen (oder Spezies oder anderer biologischer Entitäten) einer Population (oder auch einer Menge). Formal lautet die Definition: Sei $P = \{g_1, g_2, \dots, g_n\}$ eine Menge biologischer Entitäten und $D(g_i, g_j)$ ein Distanzmaß. Sei $F(d)$ die Anzahl aller Paare von Entitäten in P , die die Distanz d haben:

$$F(d) := |\{(g_i, g_j) : g_i, g_j \in P, i < j, D(g_i, g_j) = d\}|$$

Die Bedingung $i < j$ stellt sicher, daß Doppelzählungen verhindert werden und daß die Nullwerte auf der Hauptdiagonale mitgezählt werden. $f(d)$ wird nun als die relative Häufigkeit des Distanzwerts d definiert:

$$f(d) = \frac{F(d)}{\sum_i F(i)} = \frac{2 \cdot F(d)}{n \cdot (n - 1)}$$

Nunmehr kann die Distanzverteilungskomplexität definiert werden als:

$$C_D := - \sum_d f(d) \log(f(d))$$

Der Index D bezeichnet hierbei das verwendete Distanzmaß; je nach der Art der untersuchten biologischen Entitäten können durchaus mehrere verschiedene Distanzmaße für eine DVK-Analyse verwendet werden.

Die Definition der DVK enthält keine Festlegung des (zur Ermittlung der paarweisen Distanzen verwendeten) Distanzmaßes D . Schwerpunkt unserer Arbeit ist die quantitative Charakterisierung auf der molekularen Ebene. DVK-Analysen sind jedoch auch auf vielen anderen Ebenen biologischer Organisation möglich und gestatten somit vergleichende Untersuchungen der Biodiversität auf verschiedenen Organisationsebenen. Bei den im Abschnitt 6 diskutierten Analysen werden folgende Distanzmaße verwendet:

Baumpfadlänge: Dieses Distanzmaß setzt die Verfügbarkeit eines Stammbaums mit Kantenlängeninformation voraus. Die Baumpfadlänge zwischen zwei Endknoten ist die Summe der Längen aller Kanten auf dem Pfad, der die Knoten verbindet.

protdist-Distanz: `protdist` ist ein Programm zur Berechnung phylogenetischer Distanzen anhand von Proteinsequenzen (s. Abschnitt 5.2.4). Die `protdist`-Distanz ist also ein stammbaumunabhängiges Distanzmaß.

5.3.2 Diskretisierung bei realwertigen Distanzmaßen

Bei realwertigen Distanzmaßen kommt es häufig vor, daß jeder in einer Distanzmatrix vorkommende Distanzwert nur einmal in der Matrix zu finden ist. Ermittelt man bei solchen Distanzmatrizen unmittelbar die DVK, so erhält man stets den größtmöglichen Wert. Die Verteilung der einzelnen Werte, d.h. ihre ggfs. unterschiedliche Dichte, kann auf diese Weise nicht charakterisiert werden.

Man kann jedoch auch mit derartigen Distanzmaßen sinnvolle DVK-Analysen durchführen, indem man die Meßwerte vor der Ermittlung der DVK diskretisiert. Hierzu wird der Bereich zwischen dem minimalen und dem maximalen Distanzwert in gleich große Intervalle eingeteilt, und für jedes Intervall wird ermittelt, wie viele Distanzwerte d_{ij} in ihm enthalten sind. Diese Häufigkeiten werden dann für die DVK-Analyse verwendet.

Bei diesem Vorgehen ergibt sich die Anzahl der Diskretisierungsintervalle ν als zusätzlicher Parameter bei der DVK-Analyse. Die Aussagekraft der Analyse hängt in kritischer Weise von der Auswahl dieses Parameters ab. Sowohl wenn ν zu klein als auch wenn es zu groß gewählt wird, werden die Möglichkeiten zur Unterscheidung von Distanzverteilungen mit unterschiedlicher Charakteristik beeinträchtigt, oder sie gehen im Extremfall ganz verloren¹. Eine optimale Auswahl für ν sollte ein ausgewogenes Verhältnis zwischen der Anzahl der Intervalle und der mittleren Anzahl der den Intervallen zugeordneten Distanzwerten gewährleisten. Bei einer Matrix der Distanzen zwischen n Entitäten sind bei $\nu \approx n/\sqrt{2}$ der Mittelwert der Häufigkeiten und ν annähernd gleich groß. Bei $\nu = n$ ergibt sich eine mittlere Häufigkeit von $(n-1)/2$, so daß auch in diesem Fall Intervallanzahl und mittlere Häufigkeit vergleichbare Größenordnungen aufweisen.

Bei einer gründlichen DVK-Analyse sollten unterschiedliche Werte für die Anzahl der Diskretisierungsintervalle, etwa von $\nu = n/2$ bis $\nu = n$ verwendet werden und es sollte sichergestellt werden, daß innerhalb dieses Bereichs die Auswahl von ν keine kritischen Effekte für die relevanten Aspekte der Analyse mit sich bringt.

6 Untersuchung und Vergleich verschiedener Biodiversitätsmaße auf der Grundlage molekularer Daten

6.1 Material und Methoden

6.1.1 Eingabedaten

Auf der MADS *homepage* [KIM und THEISSEN 1999] steht eine Sammlung aller derzeit verfügbarer Aminosäuresequenzen von Proteinen mit einer MADS-Domäne zur Verfügung. Die MADS-Box Gene (vgl. Abschnitt 4.3.2) sind eine Familie von Transkriptionsfaktorgenen, die insbesondere bei der Entwicklung von Pflanzen eine zentrale Rolle spielen. Bei fast allen pflanzlichen Sequenzen dieser Genfamilie finden sich stromabwärts von der MADS-Box die I Region, gefolgt von der K-Box; die entsprechenden Proteindomänen werden MADS-Domäne, I-Domäne und K-Domäne genannt, die Aminosäuresequenz von MADS-Domäne, I-Domäne und K-Domäne wird im Folgenden als MIK Sequenz bezeichnet. Ein *multiple alignment* aller dieser MIK Sequenzen wird auf der MADS *homepage* zur Verfügung gestellt. Dieses *multiple alignment* wurde als Eingabe für die Berechnung einer Distanzmatrix mit dem Programm `protdist` aus dem Programmpaket PHYLIP [FELSENSTEIN 1989, FELSENSTEIN 1995]

¹Bei $\nu = 1$ ergibt sich für die DVK stets 0, beim Grenzwert $\nu \rightarrow \infty$ verliert sich der Effekt der Diskretisierung

verwendet. Bei neutral evolvierenden Sequenzen (s. Abschnitt 2.3) sind die mit `protdist` berechneten Distanzen Schätzwerte für die seit der Divergenz zweier Gene verstrichene Zeit.

Anhand dieser Distanzmatrix wurde mit dem *neighbor joining* Verfahren (PHYLIP Programm `neighbor`) ein Stammbaum rekonstruiert, wie in [THEISSEN et al. 1996, MÜNSTER et al. 1997, WINTER et al. 1999] beschrieben. Alle für die nachfolgend beschriebenen Analysen verwendeten Eingabedaten sind von der mit `protdist` berechneten Distanzmatrix bzw. von dem rekonstruierten Stammbaum abgeleitet.

6.1.2 Biodiversitätsmaße

6.1.2.1 Stammbaumbasierte Biodiversitätsmaße Folgende stammbaumbasierte Biodiversitätsmaße (vgl. auch Anhang) wurden bisher in die Untersuchungen einbezogen:

- Die Anzahl der Endknoten (also die Anzahl Sequenzen)
- Die Maße m_{II} , m_{III} , m_{IV} , m_V , m_{VI} und m_{VII} (s. Abschnitt 5.2.1)
- Die Baumlänge (s. Abschnitt 5.2.3)
- Die anhand der Baumpfadlängen berechnete DVK (s. Abschnitt 5.3).

6.1.2.2 Stammbaumunabhängige distanzbasierte Biodiversitätsmaße Auf der Grundlage der mit `protdist` erstellten Distanzmaße wurden folgende Biodiversitätsmaße untersucht:

- Die DVK der Distanzwerte, die durch das Programm `protdist` (aus dem Programmpaket PHYLIP) aus dem *alignment* der MIK-Sequenzen (vgl. MADS-homepage) berechnet wurden.
- Das gewichtete Divergenzmittel in der mit `protdist` berechneten Matrix (s. Abschnitt 5.2.4).
- Die gewichtete Divergenzdispersion in der mit `protdist` berechneten Matrix (s. Abschnitt 5.2.4).

6.1.3 Wurzelpositionsverschiebung

Die Stammbaumrekonstruktion mit *neighbor joining* ergibt unbewurzelte Bäume, d.h. das Rekonstruktionsverfahren erlaubt keine Rückschlüsse auf die Position der Wurzel innerhalb des Stammbaums. Über die Wurzelposition sind bei unbewurzelten Stammbäumen nur durch das Heranziehen zusätzlicher biologischer Information Abschätzungen möglich. Ein gängiges Verfahren dieser Art ist die Einbeziehung von *outgroups* in die Stammbaumanalyse. Derartige Abschätzungen sind jedoch oft problematisch.

Die Maße m_{II} , m_{III} , m_V und m_{VI} können nur anhand bewurzelter Stammbäume berechnet werden. Um das Ausmaß der Abhängigkeit dieser Maße von der Wurzelposition zu ermitteln, wurde daher ein Satz von Bäumen erstellt, die topologisch und hinsichtlich der Kantenlängen identisch sind und sich nur in der Positionierung der Wurzel (also der basalen Trifurkation) unterscheiden.

6.1.4 Simulation von Biodiversitätsverlusten

Zu den wichtigsten Aufgaben, zu deren Lösung Biodiversitätsmaße einen Beitrag leisten sollen, gehört die Beantwortung der Frage, wie groß der Verlust an Biodiversität ist, der durch den Verlust einer gegebenen Teilmenge von Entitäten (Genen, Lebewesen, Arten etc.) aus der Gesamtmenge verursacht wird. Solche Biodiversitätsverluste wurden auf unterschiedliche Weise simuliert.

6.1.4.1 Entfernung von Teilbäumen Durch die Durchtrennung einer Kante in einem Stammbaum kann dieser in zwei Teilbäume zerlegt werden. Alle Teilbäume dieser Art wurden von dem *neighbor joining* Stammbaum erzeugt.

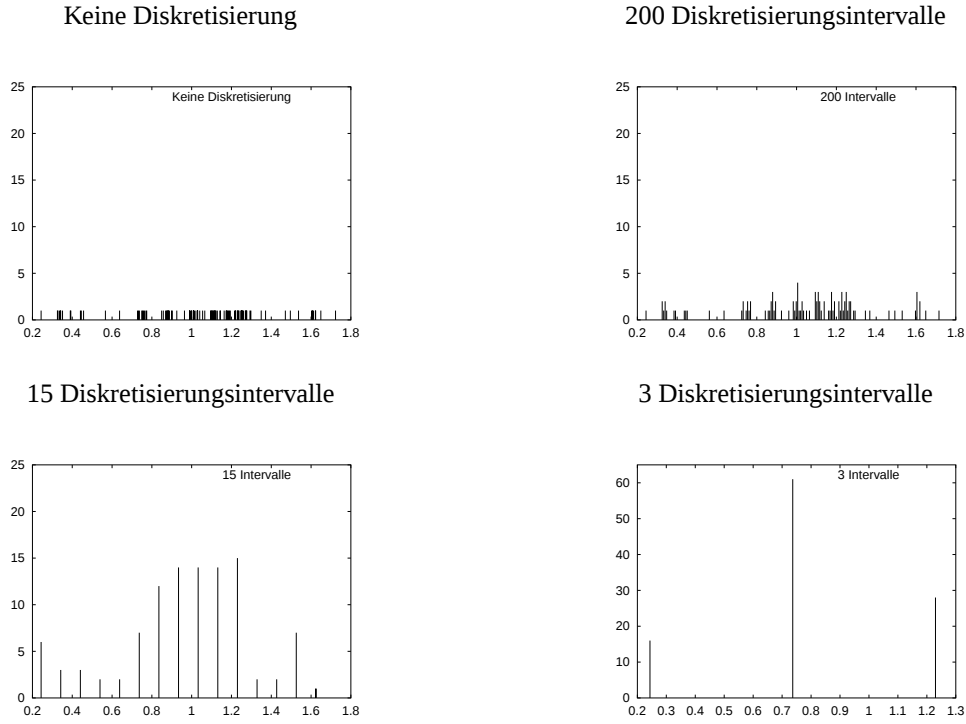


Abbildung 8: Illustration der Diskretisierung der Verteilung realwertiger Distanzwerte. Die Abbildungen zeigen Verteilungen der Baumpfadlängen in einem Stammbaum mit 15 Endknoten.

6.1.4.2 Entfernung von Teilmengen Durch das wiederholte Löschen von Entitäten können Teilmengen von Entitäten erhalten werden. Anhand einer solchen Teilmenge kann man eine verkleinerte Distanzmatrix erzeugen, indem man aus der Gesamtmatrix nur die Zeilen und Spalten, die zu in der Teilmenge enthaltenen Entitäten gehören, auswählt. Solche verkleinerten Matrizen werden im Folgenden als *RRMs* (*randomly reduced matrix*) bezeichnet.

Eine solche Teilmenge induziert weiterhin einen ebenfalls verkleinerten Stammbaum, der mit dem Gesamtstammbaum kompatibel ist [WILLSON 1998]. Diesen induzierten Stammbaum erhält man, indem man alle Endknoten, die nicht in der Teilmenge enthaltene Entitäten repräsentieren, aus dem Gesamtbaum entfernt. Derart verkleinerte Bäume werden im Folgenden als *RPTs* (*randomly pruned trees*) bezeichnet. Bei der Erstellung von *RPTs* bilden die Endknoten in der Regel keine monophyletische Clade, daher sind *RPTs* von Teilbäumen strikt zu unterscheiden.

Die Anzahl aller Teilmengen einer Gesamtmenge steigt exponentiell mit der Größe der Gesamtmenge, was eine Untersuchung aller Teilmengen von MADS-Proteinen unmöglich macht. Daher wurden zufallsgesteuert 50 Teilmengen für jede Teilmengengröße zwischen 3 und $n - 3$ für die Untersuchungen anhand von *RRMs* und *RPTs* erzeugt.

6.2 Ergebnisse und Diskussion

6.2.1 DVK-Berechnung mit unterschiedlichen Diskretisierungsintervallgrößen

Die Baumpfadlänge und die *protdist*-Distanz sind realwertige Distanzmaße; in den entsprechenden Distanzmatrizen kommen die meisten Werte nur einmal vor. Den in Abschnitt 5.3.2 dargelegten Überlegungen entsprechend müssen diese Werte diskretisiert werden, um eine informative DVK-Analyse zu ermöglichen. In diesem Abschnitt wird die Bedeutung der Diskretisierung zunächst anhand eines Beispiels illustriert und anschließend mithilfe einer Korrelationsanalyse systematisch untersucht.

Abb. 8 zeigt ein Beispiel für die Verteilung realwertiger Distanzwerte und die Diskretisierung dieser Verteilung mit unterschiedlich vielen Diskretisierungsintervallen. Ohne Diskretisierung kommt jeder Wert nur einmal

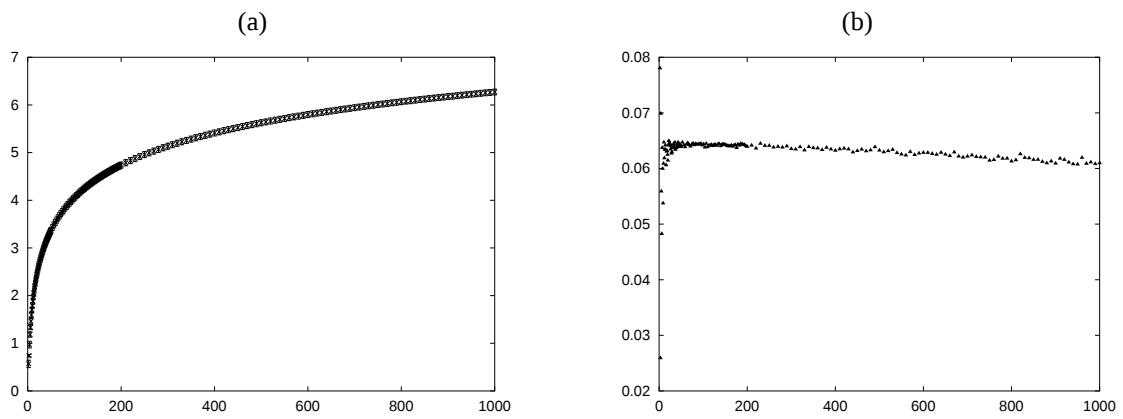


Abbildung 9: DVK-Werte (a) und Standardabweichung der DVK-Werte (b) in Abhängigkeit von der Diskretisierung.

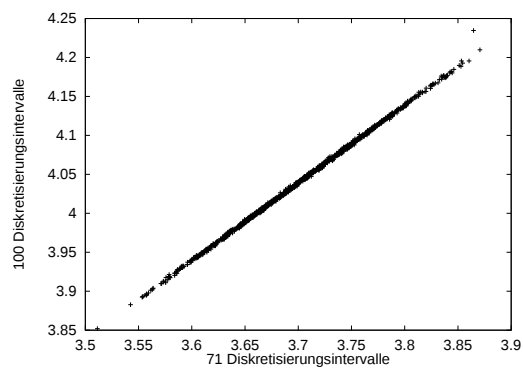


Abbildung 10: Korrelationsplot von mit $\nu = 71$ und $\nu = 100$ ermittelten DVK-Werten.

vor. Würde man hier die DVK berechnen, ergäbe sich der maximal mögliche Wert von $\log(n(n-1)/2)$. Bei einer Diskretisierung mit 200 Intervallen fallen nur wenige Distanzwerte zusammen. In der Diskretisierung mit 15 Intervallen ($n = \nu$) wird die unterschiedliche Dichte der realwertigen Distanzen gut in unterschiedliche Häufigkeiten umgesetzt. Bei 3 Intervallen ist die Auflösung dagegen deutlich zu grob, um die Charakteristik der Dichteverteilung noch widerspiegeln zu können.

Abb. 9a zeigt die Abhängigkeit der DVK von der Anzahl der Diskretisierungsintervalle. Von einem Satz aus 50 *RRMs* von jeweils 100 MADS-Proteinen wurden jeweils die Distanzverteilungen ermittelt und mit $\nu = 2, 3, \dots, 1000$ Intervallen diskretisiert. Die DVK steigt in Abhängigkeit von ν zunächst stark an. Gleichzeitig nimmt die Streuung der Werte zu. Bei größeren Werten von ν geht die Streuung wieder zurück (vgl. Abb. 9b). Das Anwachsen der DVK geht ebenfalls zurück. Bei sehr großen ν geht die DVK gegen $\log(100(100-1)/2) \approx 8.5$; sie kann diesen Maximalwert nicht übersteigen, weil es im oberen Dreieck einer Distanzmatrix $n(n-1)/2$ Distanzwerte gibt, die im Maximalfall alle unterschiedlich sein können.

Aus Abb. 9b geht hervor, daß die Streuung der DVK-Werte für die unterschiedlichen *RPTs* in Abhängigkeit von der Anzahl von Diskretisierungsintervallen ν bei sehr kleinen ν gering ist. Mit zunehmenden ν wachsen die Werte der Standardabweichung zunächst sprunghaft an. Zwischen $\nu = 65$ und $\nu = 125$ etwa treten die größten Standardabweichungen auf. Bei größeren Werten von ν fällt die Standardabweichung langsam ab. Hier ist die Feststellung interessant, daß dieser Bereich der größten Standardabweichungen den Wertebereich von $\nu = 100/\sqrt{2} \approx 70$ bis $\nu = 100$ umfaßt. Die Unterschiedlichkeit der DVK-Werte, und damit die Möglichkeit, Unterschiede zwischen *RPTs* mit gleicher Anzahl Endknoten anhand der DVK-Werte zu charakterisieren, ist also in dem in Abschnitt 5.3.2 abgeschätzten Bereich für geeignete Werte von ν am deutlichsten ausgeprägt.

Abb. 10 zeigt, daß eine extrem ausgeprägte lineare Korrelation zwischen der mit $\nu = 70$ und der mit $\nu = 100$ ermittelten DVK der *RRMs* besteht. Für andere ν zwischen 70 und 100 ergeben sich ebenfalls extrem stark ausgeprägte lineare Korrelationen.

Diese Korrelationsanalyse zeigt, daß die Auswahl der Anzahl der Diskretisierungsintervalle innerhalb des Bereichs zwischen 70 und 100 keinerlei kritische Auswirkung auf die Ergebnisse der DVK-Analyse hat. Die Untersuchung der Streuung der DVK-Werte (Abb. 9b) zeigt weiterhin, daß die DVK in diesem Wertebereich von ν eine maximal ausgeprägte Differenzierungsmöglichkeiten zwischen den gleich großen *RRMs* bietet. Diskretisierungen innerhalb dieses Wertebereiches verursachen also keine Effekte, die die Eignung der DVK zur quantitativen Charakterisierung von Biodiversität beeinträchtigen würden. Für alle im Folgenden diskutierten DVK-Messungen wurde eine Diskretisierung mit $\nu = n$ (d.h. soviele Diskretisierungsintervalle wie Entitäten) verwendet.

6.2.2 Abhängigkeit von der Wurzelposition

Abb. 11 zeigt die Auswirkung von Veränderungen der Wurzelposition auf die Maße II, III, V und VI. Die Maße IV und VII, sowie die Baumlänge und die Baumpfadlängen-Verteilungskomplexität sind wurzelpositionsunabhängig und wurden daher nicht in diese Analyse einbezogen.

Bei allen wurzelpositionsabhängigen Maßen zeigt sich in Abb. 11 zeigt ein erheblicher Einfluß der Wurzelposition auf das Meßergebnis. Bei fehlender Information über die Wurzelposition liefern Einzelmessungen mit diese Maßen also Ergebnisse, die mit erheblichen Fehlern behaftet sind. Angesichts der erheblichen Streuung der Werte kann auch die Verwendung des Mittelwerts der mit allen möglichen Wurzelpositionen erhaltenen Ergebnisse nicht als Lösung dieser Schwierigkeit angesehen werden.

Die Bestimmung der Wurzelposition bei Stammbaumrekonstruktionen anhand molekularer Daten ist jedoch stets problematisch. Dies ist letztlich darauf zurückzuführen, daß biologische Makromoleküle, welche die Grundlage für die Erhebung molekularer Daten sind, nicht über nach evolutionärem Maßstab lange Zeiträume stabil sind. Anders als z.B. morphologische Information, die in fossiler Form über mehrere hundert Jahrmillionen erhalten bleiben kann, ist biologische Information auf molekularer Ebene nur für rezente Systeme verfügbar. Dieses Problem kann nicht mithilfe technischer Mittel gelöst werden; die Wurzelpositionsabhängigkeit eines Biodiversitätsmaßes ist im Zusammenhang mit molekularen Daten daher eine Abhängigkeit von nicht verfügbarer Information. Das Ausmaß dieser Abhängigkeit, das für die Maße II, III, V und VI ermittelt wurde, schränkt die Eignung dieser Maße für die quantitative Charakterisierung von Biodiversität auf molekularer Ebene zumindest erheblich ein.

6.2.3 Charakterisierung von Biodiversitätsverlusten

6.2.3.1 Biodiversität in *RPTs* Die Auswirkungen des Verlusts von Endknoten, simuliert durch *RPTs*, ist in Abb. 12 dargestellt. Die Maße sind hier relativ zum Diversitätswert des Gesamtbaums ausgedrückt, um sie unter-

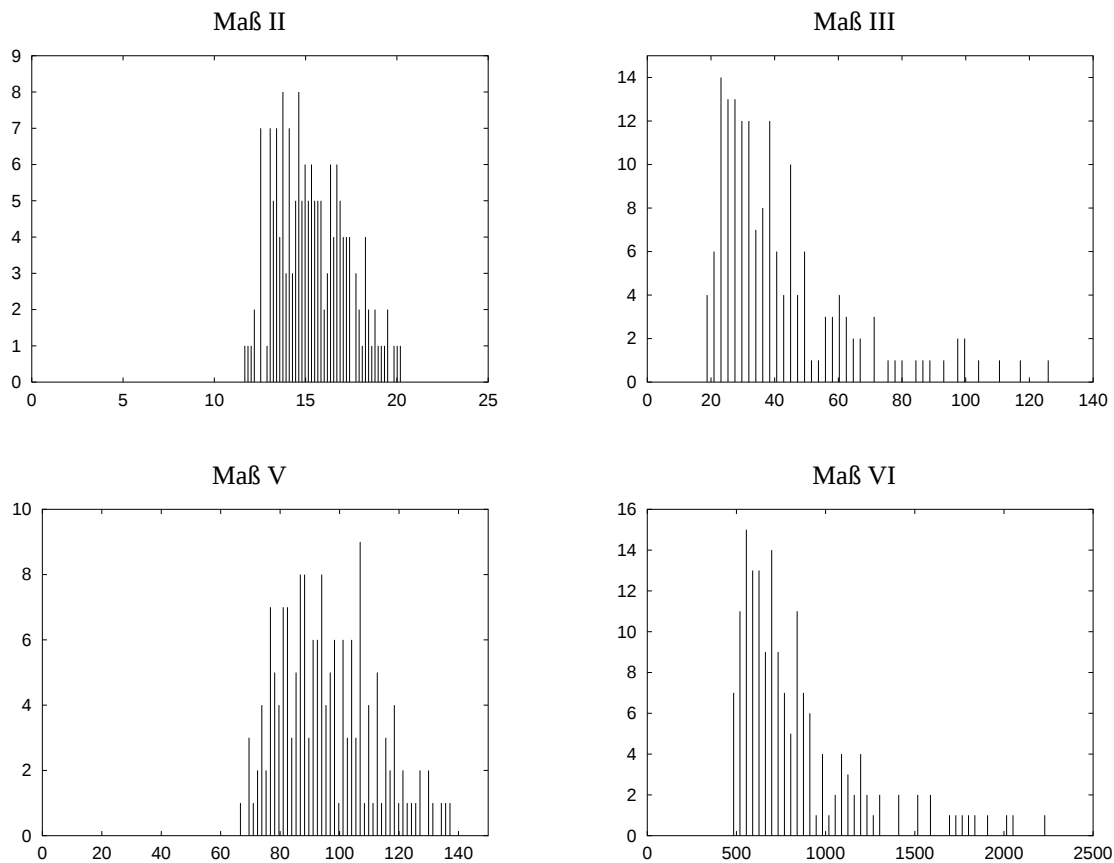


Abbildung 11: Effekt von Verschiebungen der Wurzelposition bei wurzelpositionsabhängigen Biodiversitätsmaßen. Die Werte der Maße II, III, V und VI wurden für alle möglichen Wurzelpositionen im *neighbor joining* Stammbaum der MIK-Sequenzen ermittelt. Auf der Abszisse sind die Werte des jeweiligen Biodiversitätsmaßes aufgetragen, eingeteilt in 50 Intervalle zwischen dem minimalen und dem maximalen Wert. Die Ordinate zeigt die Anzahl von Werten im jeweiligen Intervall.

einander vergleichen zu können (vgl. auch Gl. 9). Alle Maße zeigen eine stark ausgeprägte positive Korrelation mit der Anzahl der erhaltenen Endknoten.

Die Maße II bis VII, also die topologiebasierten Maße, zeigen eine lineare, positive Korrelation mit der Anzahl der im *RPT* enthaltenen Endknoten. Die stärkere Streuung bei den Maßen II, III, V und VI ist auf die Wurzelpositionsabhängigkeit dieser Maße zurückzuführen. Bei allen Analysen wurde derselbe Knoten (der durch die Ausgabereihenfolge der Knoten durch das *neighbor joining* Programm willkürlich bestimmt wurde) als Wurzelknoten angenommen. Die Lage der entfernten Endknoten im Verhältnis zu diesem kontingent bestimmten Wurzelknoten beeinflusst die Meßergebnisse bei den wurzelpositionsabhängigen Bäumen.

Bei den wurzelpositionsunabhängigen Maßen IV und VII ist die lineare Korrelation zur Anzahl der Endknoten im *RPT* besonders stark. Angesichts des Ausprägung dieser Korrelation erscheint es fraglich, inwieweit mithilfe dieser Maße ein signifikanter Erkenntnisgewinn im Vergleich zur einfachen Zählung der Anzahl der Endknoten zu erwarten ist.

Die Baumpfadlängen-VK zeigt eine nichtlineare, deutlich konvexe Korrelation mit der Anzahl der Endknoten im *RPT*. Die Korrelation ist außerdem ausgeprägt positiv, d.h. *RPTs* mit mehr Endknoten haben in den meisten Fällen eine größere Baumpfadlängen-VK. Die Baumpfadlängen-VK entspricht also der Erwartung, daß die Biodiversität in größeren Ensembles größer ist, obwohl sie die Monotonitätsforderung (s. Abschnitt 5.2.2) formal nicht erfüllt.

Ebenso wie bei der Baumpfadlängen-VK zeigt sich bei der Baumlänge eine konvexe Korrelation mit der Anzahl der Endknoten im *RPT*. Durch diese Eigenschaft unterscheiden sich diese beiden Maße qualitativ von den topologiebasierten Maßen. Nach der von EHRlich und EHRlich formulierten *rivet popper* Hypothese (zitiert

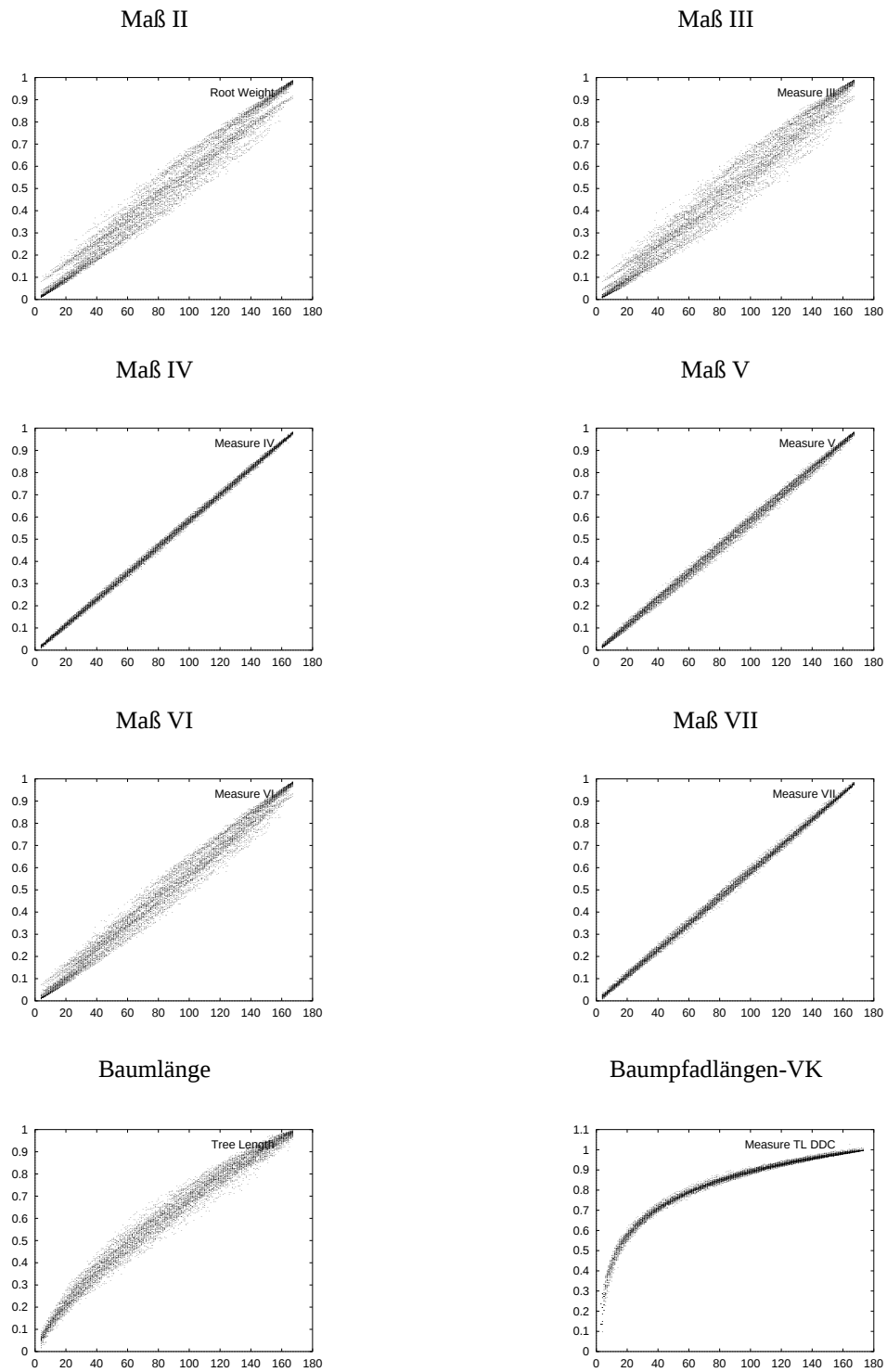


Abbildung 12: Relative Biodiversität von RPTs bei verschiedenen Biodiversitätsmaßen. Die Abszisse zeigt die Anzahl der Endknoten der RPTs, die Ordinate zeigt den Wert des jeweiligen Diversitätsmaßes relativ zum Diversitätswert des vollständigen Stammbaums.

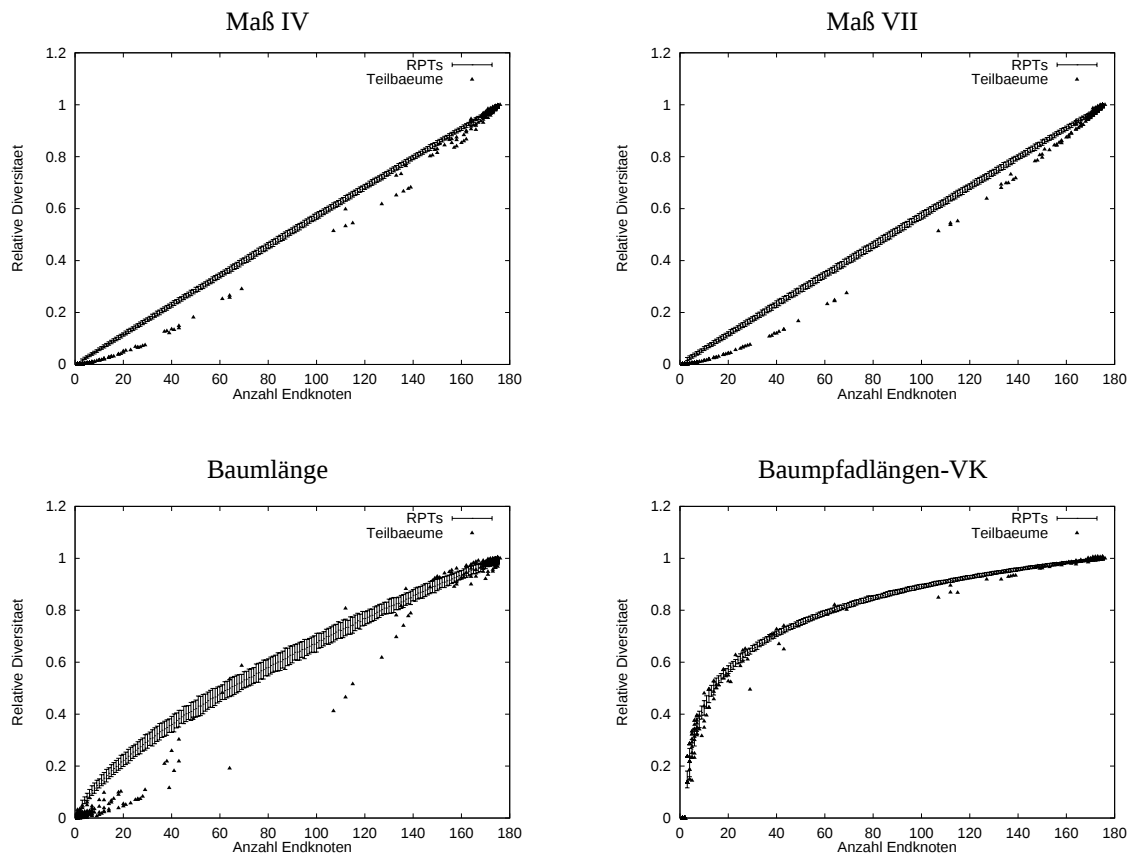


Abbildung 13: Relative Biodiversität von Teilbäumen bei verschiedenen Biodiversitätsmaßen. Die Abszisse zeigt die Anzahl der erhaltenen Endknoten der Stammbäume, die Ordinate zeigt den relativen Wert des jeweiligen Diversitätsmaßes. Der Referenzwert ist der Diversitätswert des Originalbaums. Die Biodiversitätswerte der Teilbäume sind jeweils als Punkte dargestellt, die Balken zeigen den Mittelwert und die Standardabweichung aus der Analyse der RPTs.

z.B. in [MARTINEZ 1996, S. 116]) verhält sich ein System mit Biodiversität (auf der Speziesebene) beim Verlust von Spezies so wie ein Flugzeug beim Verlust von Nieten, die das Flugzeug zusammenhalten: Während der Verlust einiger Nieten die Eigenschaften des Flugzeugs nicht wesentlich verändern, kommt es bei der Entfernung vieler Nieten zu grundlegenden (und katastrophalen) Veränderungen der Systemeigenschaften. Dieser Zusammenhang kann durch Maße, die eine konvexe, nichtlineare Korrelation mit der Anzahl erhaltener Endknoten aufweisen, möglicherweise erfasst werden, während bei Maßen mit linearer Korrelation feststeht, daß sie diese wesentliche systemare Charakteristik der Biodiversität nicht erfassen können.

6.2.3.2 Biodiversität in Teilbäumen Abb. 13 zeigt die Effekte des Entfernens von Teilbäumen auf die Werte der wurzelpositionsunabhängigen, stammbaumbasierten Biodiversitätsmaße. Für die wurzelpositionsabhängigen Maße ist diese Analyse nicht in aussagefähiger Weise durchführbar, weil die Effekte der willkürlichen Positionierung der Wurzel einen zu starken Einfluß auf die Ergebnisse haben.

Die Diagramme für die Maße IV und VII sind einander sehr ähnlich, in beiden liegen die Meßwerte für Teilbäume signifikant unterhalb der Werte für RPTs. Dies ist auf die kürzeren mittleren Distanzen in den durch Abschneiden von Teilbäumen verkleinerten Bäumen zu erklären. Die Maße IV und VII können also zwischen dem gleichmäßig gestreuten Verlust von Entitäten und dem Verlust zusammenhängender Claden unterscheiden und charakterisieren den Verlust von Claden als signifikant höher.

Bei der Baumlänge liegt der größte Teil der Meßwerte für Teilbäume ebenfalls unterhalb derjenigen für RPTs, die Differenz ist jedoch erheblich stärker ausgeprägt als bei den Maßen IV und VII. Die Baumlängen der Teilbäume sind außerdem sehr viel stärker gestreut als die Werte der Maße IV und VII. Auffallenderweise haben Teilbäume

Längen, die signifikant über der mittleren Baumlänge von *RPTs* mit gleicher Anzahl Endknoten liegen. Bei diesen Teilbäumen handelt es sich um Claden mit vielen besonders langen Kanten.

Die Baumlänge wurde als Biodiversitätsmaß, das die durch eine Speziesmenge repräsentierte *evolutionary history* quantifiziert, vorgeschlagen. Beim Verlust einer zusammenhängenden Clade geht mehr evolutionsgeschichtliche Information verloren, weil nicht nur die Information über die ausgestorbenen Spezies selbst, sondern auch die Information über deren gemeinsame Vorfahren verlorengeht. Im Stammbaum führt dies zum Wegfall aller Kanten zwischen den internen Knoten des entfernten Teilbaums. Dementsprechend resultiert das Herausschneiden eines Teilbaums in den meisten Fällen in einem stärkeren Rückgang der Baumlänge als der Verlust entsprechend vieler, über verschiedene Claden verteilter Spezies.

In molekularen Phylogenien findet man jedoch häufig Claden, in denen längere Kanten gehäuft auftreten. In diesen Claden hat offenbar eine beschleunigte evolutionäre Diversifizierung stattgefunden. Dies kann auf die Evolution neuer Funktionalität zurückgehen und insofern durchaus ein Vorgang sein, der die Biodiversität überdurchschnittlich vergrößert. Die überdurchschnittliche Länge solcher Teilbäume kann also u.U. eine erhöhte Biodiversität repräsentieren. Die Selektionseinflüsse, welche für eine biologische Strukturierung von Diversität erforderlich sind, werden durch die Messung der Baumlänge jedoch nicht erfaßt.

Die Länge von Stammbäumen kann also als eine Art Komplexitätsmaß 1. Ordnung angesehen werden, mit dem die Diversität in Bäumen und Teilbäumen charakterisiert werden kann. Zur Unterscheidung zwischen strukturierter Biodiversität und unstrukturierter Diversität ist die Baumlänge jedoch nicht geeignet.

Bei der Baumpfadlängen-VK unterscheiden sich die Meßwerte für die meisten Teilbäume nicht signifikant von den Werten von *RPTs* mit gleicher Anzahl Endknoten. Es gibt jedoch einige Teilbäume, deren Baumpfadlängen-VK signifikant kleiner als die Baumpfadlängen-VK gleich großer *RPTs* ist. Die Baumpfadlängen-VK ist also zur Unterscheidung von Teilbäumen und *RPTs* nur bedingt geeignet, während sie eine Differenzierung von Teilbäumen gestattet.

Zusammenfassend ist festzustellen, daß die hier untersuchten Maße auf den Unterschied zwischen Teilbäumen und *RPTs* qualitativ unterschiedlich reagieren. Um zu entscheiden, welches der Maße am besten zur Charakterisierung biodiversitätsrelevanter Unterschiede zwischen Teilbäumen und *RPTs* geeignet ist, sind weitergehende Untersuchungen erforderlich. Hierfür würde sich z.B. die Bestimmung der evolutionären Potentiale von Teilbäumen und *RPTs* anhand von Computersimulationen der Evolution anbieten.

6.2.3.3 Biodiversität in *RRMs* Analog zur Erstellung von *RPTs* aus einem Gesamtbaum kann der Verlust von Entitäten bei Distanzmatrizen durch die Erstellung von *RRMs* (s. Abschnitt 6.1.4) simuliert werden. Abb. 14 zeigt die Ergebnisse einer solchen Analyse. Das gewichtete Divergenzmittel und die gewichtete Divergenzdispersion zeigen eine lineare, positive Korrelation mit der Anzahl der Entitäten in der *RRM*. Bei der DVK ist dagegen eine konvexe, positive Korrelation festzustellen, die der DVK der Baumpfadlängen sehr ähnlich ist.

Der lineare Zusammenhang zwischen gewichtetem Divergenzmittel und der Anzahl Entitäten sowie zwischen gewichteter Divergenzdispersion und der Anzahl Entitäten kann bei den *RRMs* recht einfach formal erklärt werden. Die bei der Erstellung einer *RRM* ausgewählte Teilmenge T aus der Gesamtmenge G von Entitäten ist eine Zufallsstichprobe. Die Erwartungswerte für den Mittelwert und die Standardabweichung der $d(x, y)$ für $x, y \in T$ unterscheiden sich also nicht von denen der $d(x, y)$ für $x, y \in G$. Somit wird nur der erste Faktor in Gl. 7, nämlich n , durch das Ziehen der Zufallsstichprobe beim Erstellen einer *RRM* in gerichteter Weise verändert, während der Rest des Terms, der Mittelwert der $d(x, y)$, nur in stochastischer, ungerichteter Weise verändert wird.

Diese Betrachtung zeigt, daß alle auf dem Konzept des mit der Entitätenanzahl gewichteten Distanz- oder Divergenzmittels basierenden Biodiversitätsmaße eine lineare Korrelation mit der Anzahl der Entitäten aufweisen. Diese Eigenschaft gilt auch für die stammbaumbasierten Biodiversitätsmaße II bis VII, weil hier auch bei der Messung der Biodiversität von Teilmengen der Stammbaum der Gesamtmenge verwendet wird (s. Abschnitt 5.2.1), das Erstellen von Teilmengen läßt sich daher auf das Erstellen einer *RRM* aus der Matrix der $d(x, y)$ zurückführen.

Die Ähnlichkeit der Korrelation der DVK zur Größe der *RRM* mit der Korrelation der Baumpfadlängen-VK mit der Anzahl der Endknoten des *RPT* erklärt sich daraus, daß die distanzbasierten Stammbaumrekonstruktionsverfahren, darunter auch *neighbor joining*, so gestaltet sind, daß die Differenzen zwischen der Eingabematrix und der Matrix der Baumpfadlängen minimiert werden – die Ähnlichkeit dieser Matrizen dient letztlich dazu, die Übereinstimmung des rekonstruierten Stammbaums mit den empirisch ermittelten Eingabedaten zu bestimmen.

Beim Vergleich der beiden Korrelationsplots fällt jedoch auf, daß die Streuung der DVK bei den *prot-dist-RRMs* stärker ist als bei den Baumpfadlängen. Dies ist möglicherweise ein Hinweis darauf, daß bei der Stammbaumrekonstruktion mit *neighbor joining* Signale, die die Unterscheidung gleich großer Teilmengen anhand der DVK

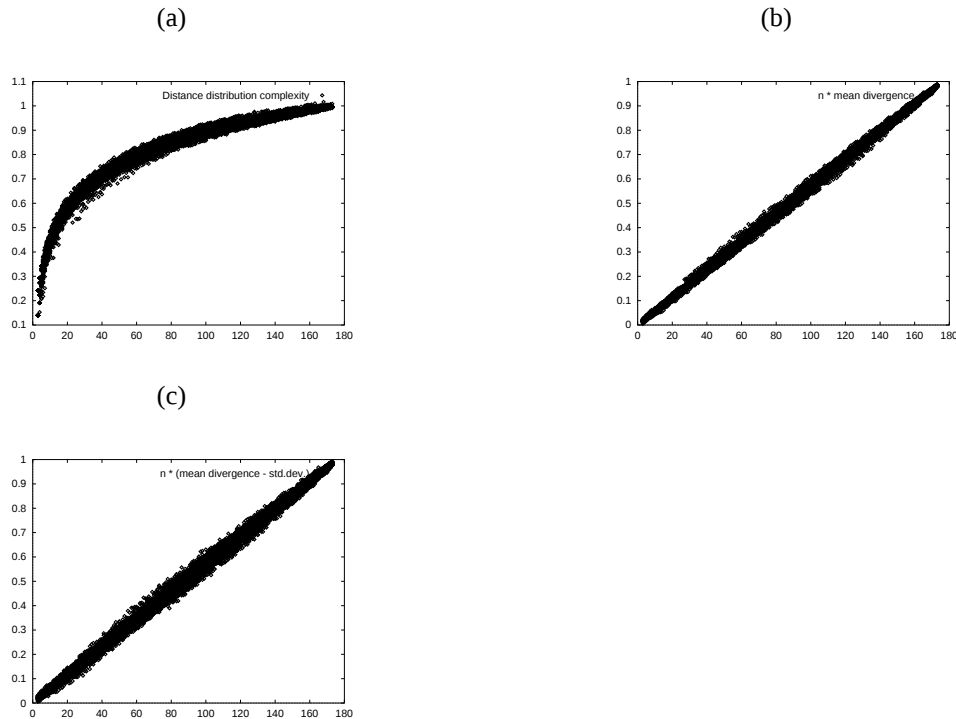


Abbildung 14: Relative Biodiversität in RRM bei verschiedenen Biodiversitätsmaßen: (a) DVK, (b) gewichtetes Divergenzmittel, (c) gewichtete Divergenzdispersion. Die Abszisse zeigt die Anzahl der erhaltenen Zeilen und Spalten in der Matrix, die Ordinate zeigt den relativen Wert des jeweiligen Diversitätsmaßes. Der Referenzwert ist der mit der vollständigen Matrix ermittelte Diversitätswert.

ermöglichen, abgeschwächt werden. DVK-Analysen auf der Basis genetischer Distanzen sind in diesem Fall nicht nur stammbaumbasierten DVK-Analysen vorzuziehen, weil der Rekonstruktionsprozeß mit verschiedenen Problemen behaftet sein kann, sondern auch deshalb, weil der Rekonstruktionsschritt mit einem unnötigen Verlust biodiversitätsrelevanter Information verbunden ist.

6.3 Zusammenfassende Diskussion

Verschiedene Autoren haben seit Beginn der neunziger Jahre versucht, Methoden und Maße zu entwickeln, um Biodiversität adäquater und präziser zu charakterisieren, als dies anhand der Zählung von Arten möglich ist. Verglichen mit Artenzahlen erfordert die Berechnung solcher Maße mehr Information.

Mit unserer Studie legen wir eine vergleichende Untersuchung verschiedener vorgeschlagener Biodiversitätsmaße vor. In der Literatur fehlen solche Vergleiche bisher fast völlig. Unsere Ergebnisse lassen sich wie folgt zusammenfassen:

- Die in [WILLIAMS et al. 1991] vorgeschlagenen, ausschließlich auf Grundlage der Topologie eines Stammbaums berechneten Maße haben sich als sehr stark linear abhängig von der Anzahl der Entitäten im betrachteten Ensemble erwiesen. Dies steht im Widerspruch zu dem Anspruch, mithilfe dieser Maße Biodiversität in einer Weise zu erfassen, die über das einfache Artenzählen hinausgeht.
- Distanz- oder divergenzbasierte Biodiversitätsmaße bieten einen von Artkonzepten unabhängigen Zugang zur quantitativen Charakterisierung von Biodiversität. Voraussetzung hierfür ist, daß der Distanz- oder Divergenzwert zwischen zwei Entitäten unabhängig davon ist, ob die betrachteten Entitäten derselben Art (oder einer anderen taxonomischen Kategorie) zugeordnet werden oder nicht.
- Obwohl die DVK die Monotonitätsforderung formal nicht erfüllt, stehen die in unserer Studie erhaltenen Meßergebnisse nicht in Widerspruch zu der „intuitiven“ Erwartung höherer Biodiversitätswerte bei größeren Entitätenzahlen, aufgrund derer die Monotonitätsforderung vorgeschlagen wurde.

- Die DVK zeigt in Abhängigkeit von der Anzahl erhaltener Entitäten ein konvexes Verhalten, d.h. der (simulierte) Verlust vieler Entitäten führt im Vergleich zum Verlust weniger Entitäten zu einem überproportionalen Rückgang der DVK. In weniger ausgeprägter Weise gilt dies auch für die Baumlänge. Dieses Verhalten entspricht qualitativ der Bedeutung, die zumindest bestimmte Formen der Biodiversität für die Stabilität eines Ökosystems haben. Insofern entspricht diese qualitative Eigenschaft den Erwartungen an ein Maß, das Biodiversität adäquater als die Artenzahl beschreibt.

Die ausgeprägte lineare Korrelation zwischen der Artenzahl und den auf Stammbaumtopologien basierten Biodiversitätsmaßen konnte auf das formale Grundmuster der Definition dieser Maße zurückgeführt werden (s. Abschnitt 6.2.3.3). Zu dem Ansatz, Biodiversität anhand der Topologie von Stammbäumen zu charakterisieren, ist weiterhin anzumerken, daß die internen Knoten in einem Stammbaum nur eine kleine, willkürliche Auswahl aus der Gesamtheit der Vorfahren der betrachteten rezenten Entitäten darstellt. Alle Entitäten, die die Abstammungslinien zwischen den durch interne Knoten oder Endknoten repräsentierten Entitäten bilden, werden in der elementaren Stammbaumtopologie nicht repräsentiert. Die Auswahl der durch interne Knoten repräsentierten Entitäten ist also artifiziell. Dies Problem wird dadurch, daß diese Auswahl in kritischer Weise von der Menge der betrachteten rezenten Entitäten, also vom *sampling*, abhängt, weiter verschärft. Durch die Einbeziehung von Kantenlängeninformation, die z.B. die Anzahl der Entitäten in einer durch eine Kante repräsentierten Abstammungslinie repräsentieren kann, kann dieses Problem der *sampling*-Abhängigkeit zumindest teilweise gelöst werden.

Molekulargenetische Daten werden in zunehmendem Umfang als Grundlage zur Entwicklung der Taxonomie von Pflanzen genutzt (s. [NYFFELER 1999]). Daher ist es sinnvoll, auch die Charakterisierung von Biodiversität auf der Basis molekularbiologischer Daten vorzunehmen. In Verbindung mit distanz- oder divergenzbasierten Ansätzen ermöglichen solche Daten theoretisch eine von artkonzeptbedingten Verzerrungen freie Charakterisierung von Biodiversität. In der Praxis wurde anhand molekularbiologischer Information die Taxonomie der Angiospermen gerade in jüngerer Zeit reevaluiert [MATHEWS und DONOGHUE 1999, PARKINSON et al. 1999, QIU et al. 1999, SOLTIS et al. 1999]. Diese Untersuchungen bestätigten die Auffassungen der organismischen Taxonomie in den meisten Fällen, aber sie führten auch zu einigen Korrekturen [WINTER et al. 1999]. Die in diesem Beitrag anhand der MADS-Box-Genfamilie vorgestellten Methoden zur Charakterisierung von Biodiversität anhand molekularer Daten sollten daher in Zukunft auf der Grundlage weiterer Sammlungen molekularer Daten, wie sie z.B. in [MATHEWS und DONOGHUE 1999, PARKINSON et al. 1999, QIU et al. 1999, SOLTIS et al. 1999, WINTER et al. 1999] zusammengestellt wurden, weitergehend erprobt und etabliert werden.

Trotz allen Anwachsens der Menge verfügbarer molekularer Information verbleibt meßpraktisch jedoch das Problem, daß die Erhebung und Auswahl des molekulargenetischen Datenmaterials durch Artkonzepte beeinflusst werden kann. Ausmaß und Art der hierdurch hervorgerufenen Störungen sollten anhand statistischer Analysen und von Simulationsexperimenten näher untersucht werden.

Die Charakterisierung von DVK und Baumlänge als Maße, die die ökologische Bedeutung der Biodiversität zumindest in einiger Hinsicht adäquater als die Artenzahl erfassen, zeigt an, daß sowohl stammbaumbasierte als auch stammbaumunabhängige, distanzbasierte Biodiversitätsmaße die Möglichkeit bieten, durch Einbeziehung von Information über die Beziehungen biotischer Entitäten untereinander zu Biodiversitätsmaßen mit qualitativ neuen Eigenschaften zu gelangen. Diese neuen Eigenschaften, insbesondere mögliche Bezüge zu ökologischen und evolutionären Prozessen, sollten weitergehend untersucht werden, u.a. deshalb, weil solche Untersuchungen auch zu weiterführenden Erkenntnissen über die Biodiversität führen können.

7 Zusammenfassung

„Biodiversität“ ist zu einem überaus populären Begriff geworden, der nicht nur innerhalb der Naturwissenschaften, sondern auch in vielen anderen Bereichen von Gesellschaft und Politik verwendet wird. In bemerkenswertem Gegensatz zur Gebrauchshäufigkeit des Begriffes steht allerdings die unscharfe Definition von Biodiversität. Oftmals wird die Anzahl der Arten in einem bestimmten Areal als Maß der Biodiversität genommen, doch dies ist aus mehreren Gründen unbefriedigend. So gibt es in der Biologie mehrere Artdefinitionen, deren Anwendung zu unterschiedlichen Artenzahlen führt. Weiterhin gibt es bisher keinen Artbegriff, der sich in einheitlicher Weise auf alle Organismen anwenden läßt. Darüberhinaus bleibt bei einer schlichten Erfassung der Artenzahl die Verschiedenheit der Arten unberücksichtigt. Und schlußendlich existiert Biodiversität auf allen Ebenen biologischer Organisation, von der genomischen Ebene bis hin zum Niveau der Ökosysteme. Biodiversität außerhalb der Artebene wird aber von artbasierten Biodiversitätsmaßen nicht erfaßt, obwohl auch hier interessante Phänomene zu beschreiben sind.

Um dies zu verdeutlichen, haben wir am Beispiel der Transposons erläutert, wie Diversität auf genomischer Ebene erzeugt und strukturiert werden kann. Am Beispiel des Kulturmais und seines wilden Vorfahrens, der Teosinte, haben wir gezeigt, daß Diversität auf genomischer und morphologischer Ebene stark voneinander abweichen können, weil einzelne Entwicklungskontrollgene einen weit überproportionalen Beitrag zur morphologischen Diversität leisten können. Teosinte und Mais sind darüberhinaus ein gutes Beispiel für drastische morphologische Diversität innerhalb einer Art.

Um solche Phänomene adäquat beschreiben zu können, müssen Biodiversitätsmaße entwickelt werden, die auf allen Ebenen biologischer Organisation und auf alle Organismen angewendet werden können. Als Ausgangspunkt bieten sich Distanzmaße an, die die Unterschiedlichkeit von Objekten quantitativ beschreiben. Dies ist insbesondere auf der Ebene von Nukleinsäure- und Proteinsequenzen leicht möglich. Genombasierte Biodiversitätsmaße haben zudem den Vorteil, bei allen Lebewesen ermittelt werden zu können. Sie besitzen überdies das Potential, durch standardisierte und automatisierte Verfahren mit hohem Probendurchsatz bestimmbar zu sein. Dies erst eröffnet die Möglichkeit, eine Totalaufnahme von Biodiversität durchführen zu können, bei der jedes Individuum quantitativ berücksichtigt wird. Ein besonderer Vorteil derartiger, nicht taxonomiebasierter Verfahren besteht darin, daß keine Verzerrungen bei der Probennahme in einem Biotop durch die Einordnung in Taxa entstehen können.

Aus der Literatur waren unterschiedliche Biodiversitätsmaße auf der Grundlage von Distanzmaßen verfügbar, die in unserem Artikel vergleichend untersucht wurden. Aus Distanzwerten lassen sich recht leicht evolutionäre Stammbäume konstruieren. Ein daraus abgeleitetes Maß für Biodiversität ist die Baumlänge, d.h. die Summe der Längen aller Kanten des Stammbaumes. Ein solches recht einfaches Biodiversitätsmaß hat unter anderem den Vorteil, daß es sowohl die Anzahl, als auch die Verschiedenheit der biologischen Einheiten (z.B. Spezies) berücksichtigt. Es besitzt darüberhinaus einige formale Eigenschaften, die für Biodiversitätsmaße vorgeschlagen wurden, z.B. erfüllt es die Monotonitätsforderung. Nach dieser soll die Diversität einer Teilmenge von biologischen Einheiten kleiner oder höchstens gleich der Diversität der Gesamtmenge der Einheiten sein.

Die Baumlänge und die meisten anderen Biodiversitätsmaße haben allerdings den Nachteil, daß sie nicht zwischen unstrukturierter und strukturierter Diversität unterscheiden. Unstrukturierte Diversität kann z.B. durch reine Zufallsprozesse entstehen, während durch biologische Evolutionsprozesse entstandene Diversität (d.h. Biodiversität im engeren Sinne) stets strukturiert ist. Um diesen Aspekt der Biodiversität zu erfassen, bietet sich die Distanzverteilungskomplexität (DVK) an. Diese ist als die Shannon-Entropie der Verteilung der paarweisen Distanzen der biologischen Einheiten definiert. Die DVK erfüllt allerdings die Monotonitätsforderung nicht, und es ist daher umstritten, ob sie sich als alleiniges Biodiversitätsmaß eignet. Eine adäquate Charakterisierung von Biodiversität könnte die Verwendung mehrerer Maße, z.B. Baumlänge und DVK, erfordern, was jedoch weiterer theoretischer Untersuchungen bedarf.

Zur praktischen Anwendung der hier dargelegten Ideen müssen desweiteren die technischen Voraussetzungen geschaffen werden, um Distanzen biologischer Objekte in großem Stil kostengünstig messen zu können. Dies könnte auf der Nukleinsäureebene durch den Einsatz modernster Methoden der Genomanalyse geschehen, z.B. unter Verwendung von „DNA-Chips“. Die Adaptation solcher Verfahren zur Messung von Biodiversität sollte entschlossen vorangetrieben werden. Wir schlagen vor, daß zu diesem Zweck ein interdisziplinäres Zentrum für Biodiversitätsforschung etabliert wird, welches sich sowohl mit den theoretischen, als auch den praktischen Aspekten der Biodiversität beschäftigt.

Literatur

- [AVISE und WOLLENBERG 1997] AVISE, J.C. und K. WOLLENBERG (1997). *Phylogenetics and the origin of species*. PNAS, 94:7748–7755.
- [BACHMANN 1998] BACHMANN, KONRAD (1998). *Species as Units of Diversity: An Outdated Concept*. Theory Biosci., 117:213–230.
- [BANKS und FEDOROFF 1989] BANKS, J.A. und N. FEDOROFF (1989). *Patterns of Developmental and Heritable Change in Methylation of the Suppressor-Mutator Transposable Element*. Developmental Genetics, 10:425–437.
- [BARRATT et al. 1997] BARRATT, E. M., R. DEAVILLE, T. M. BURLAND, G. JONES, P. A. RACEY und R. K. WAYNE (1997). *DNA answers the call of pipistrelle bat species*. Nature, 387:138–139.

- [BEADLE 1980] BEADLE, G.W. (1980). *The ancestry of corn*. Sci.Am., 242:96–103.
- [BOEKE und DEVINE 1998] BOEKE, JEF D. und S. E. DEVINE (1998). *Yeast Retrotransposons: Finding a Nice Quiet Neighborhood*. Cell, 93:1087–1089.
- [BRADLEY et al. 1993] BRADLEY, D., R. CARPENTER, H. SOMMER, N. HARTLEY und E. COEN (1993). *Complementary Floral Homeotic Phenotypes Result From Opposite Orientations of a Transposon at the plena Locus of Antirrhinum*. Cell, 22:85–95.
- [BUCKLER und HOLTSFORD 1996] BUCKLER, E.S. IV und T. HOLTSFORD (1996). *Zea systematics: ribosomal ITS evidence*. Mol.Biol.Evol, 13:612–622.
- [CAPORAL 1999] CAPORAL, LYNN HELENA, Hrsg. (1999). *Molecular Strategies in Biological Evolution*, Bd. 870 d. Reihe *Annals of the New York Academy of Sciences*. New York Ac Sc, New York.
- [COLOMBO et al. 1995] COLOMBO, L., J. FRANKEN, E. KOETJE, J. VAN WENT, H. DONS, G. ANGENENT und A. VAN TUNEN (1995). *The petunia MADS box gene FBP11 determines ovule identity*. Plant Cell, 7:1859–1868.
- [CRAWFORD 1990] CRAWFORD, DANIEL J. (1990). *Plant Molecular Systematics*. John Wiley & Sons, New York.
- [CROZIER und KUSMIERSKI 1994] CROZIER, R.H. und R. KUSMIERSKI (1994). *Genetic Distances and the Setting of Conservation Priorities*. In: LOESCHCKE, V., J. TOMIUK und J. KAIN, Hrsg.: *Conservation Genetics*, S. 1311–1323, Basel, Switzerland. Birkhäuser Verlag.
- [DAVIES und SCHWARZ-SOMMER 1994] DAVIES, B. und Z. SCHWARZ-SOMMER (1994). *Control of floral organ identity by homeotic MADS-box transcription factors*. In: NOVER, L., Hrsg.: *Results and problems in cell differentiation*, Bd. 20, S. 235–258, Berlin. Springer Verlag.
- [DOEBLEY 1990a] DOEBLEY, J. (1990a). *Molecular evidence and the evolution of maize*. Economic Bot., 3 Suppl.:6–27.
- [DOEBLEY 1990b] DOEBLEY, J. (1990b). *Molecular Systematics of Zea (Gramineae)*. Maydica, 35:143–150.
- [DOEBLEY 1992] DOEBLEY, J. (1992). *Mapping the genes that made maize*. Trends in Genet., 8:302–307.
- [DOEBLEY et al. 1997] DOEBLEY, J., A. STEC und L. HUBBARD (1997). *The evolution of apical dominance in maize*. Nature, 386:485–488.
- [DUBREUIL und CHARCOSSET 1998] DUBREUIL, P. und A. CHARCOSSET (1998). *Genetic Diversity Within and Among Maize Populations: A Comparison between Isozyme and Nuclear RFLP Loci*. Theor Appl Genet, 96:577–587.
- [EHRlich und WILSON 1991] EHRlich, PAUL R. und E. O. WILSON (1991). *Biodiversity Studies: Science and Policy*. Science, 253:758–762.
- [FELSENSTEIN 1989] FELSENSTEIN, J. (1989). *PHYLIP – Phylogeny Inference Package (Version 3.2)*. Cladistics, 5:164–166.
- [FELSENSTEIN 1995] FELSENSTEIN, J. (1995). *PHYLIP – Phylogeny Inference Package 3.5c*. Distributed by the author; available via ftp from <ftp.genetics.washington.edu>.
- [HAGEMANN 1999] HAGEMANN, RUDOLF (1999). *Allgemeine Genetik*. Urban und Fischer, München.
- [HANSON et al. 1999] HANSON, BROOKS, G. CHIN, A. SUGDEN und E. CULOTTA (1999). *The Diversity of Evolution*. Science, 284:2105.
- [HU et al. 1975] HU, SYLVIA, E. OKTSUBO, N. DAVIDSON und H. SAEDLER (1975). *Electron Microscope Heteroduplex Studies of Sequence Relations Among Bacterial Plasmids: Identification and Mapping of the Insertion Sequences IS1 and IS2 in F and R Plasmids*. J. Bacteriology, 122:764–775.

- [HUIJSER et al. 1992] HUIJSER, P., J. KLEIN, W.-E. LÖNNIG, H. MEIJER, H. SAEDLER und H. SOMMER (1992). *Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene squamosa in Antirrhinum majus*. EMBO J., 11:1239–1249.
- [KIDWELL und LISCH 1997] KIDWELL, MARGARET G. und D. LISCH (1997). *Transposable Elements as Sources of Variation in Animals and Plants*. PNAS, 94:7704–7711.
- [KIM 1996a] KIM, JAN T. (1996a). *Distance Distribution Complexity: A Measure for the Structured Diversity in Evolving Populations..* In: LANGTON, CHRISTOPHER G. und K. SHIMOHARA, Hrsg.: *Artificial Life V*, S. 281–288, Cambridge, MA. MIT Press.
- [KIM 1996b] KIM, JAN T. (1996b). *Untersuchungen zur Evolution von morphologischer und taxonomischer Diversität und Komplexität anhand von Computermodellen..* Doktorarbeit, Universität zu Köln. <ftp://ftp.mpiz-koeln.mpg.de/pub/mpiz/zwdv/kim/diss/kimdiss.ps.gz>.
- [KIM und THEISSEN 1999] KIM, JAN T. und G. THEISSEN (1999). *The MADS-box Gene Home Page*. <http://www.mpiz-koeln.mpg.de/mads/>.
- [KNIPPERS 1997] KNIPPERS, ROLF (1997). *Molekulare Genetik*. Thieme, Stuttgart.
- [KUNZE et al. 1997] KUNZE, R., H. SAEDLER und W.-E. LÖNNIG (1997). *Plant Transposable Elements*. Advances in Botanical Research, 27:331–470.
- [KURTHS et al. 1995] KURTHS, J., U. SCHWARZ, A. WITT, R. KRAMPE und M. ABEL (1995). *Measures of Complexity in Signal Analysis*. In: KATZ, RICHARD A., Hrsg.: *Chaotic, Fractal, and Nonlinear Signal Processing*, S. 33–54, New York. Woodbury.
- [LEBRUN et al. 1998] LEBRUN, P., Y. N'CHO, M. SEGUIN, L. GRIVET und L. BAUDOUIN (1998). *Genetic Diversity in Coconut (Cocos nucifera L.) Revealed by Restriction Fragment Length Polymorphism (RFLP) Markers*. Euphytica, 101:103–108.
- [LI und GRAUR 1991] LI, WEN-HSIUNG und D. GRAUR (1991). *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, Mass.
- [LIVINI et al. 1992] LIVINI, C., P. AJMONE-MARSAN, A. MELCHINGER, M. MESSMER und M. MOTTO (1992). *Genetic Diversity of Maize Inbred Lines Within and Among Heterotic Groups Revealed by RFLPs*. Theor Appl Genet, 84:17–25.
- [MANDEL et al. 1992] MANDEL, M.A., C. GUSTAFSON-BROWN, B. SAVIDGE und M. YANOFSKY (1992). *Molecular characterization of the Arabidopsis floral homeotic gene APETALA1*. Nature, 360:273–277.
- [MARTINEZ 1996] MARTINEZ, NEO D. (1996). *Defining and Measuring Functional Aspects of Biodiversity*. In: GASTON, KEVIN, Hrsg.: *Biodiversity - A Biology of Numbers and Difference*, S. 114–148. Blackwell Science Ltd, London.
- [MATHEWS und DONOGHUE 1999] MATHEWS, SARAH und M. J. DONOGHUE (1999). *The Root of Angiosperm Phylogeny Inferred from Duplicate Phytochrome Genes*. Science, 286:947–950.
- [MCGRADY-STEED et al. 1997] MCGRADY-STEED, JILL, P. M. HARRIS und P. J. MORIN (1997). *Biodiversity Regulates Ecosystem Predictability*. Nature, 390:162–165.
- [MEYEROWITZ 1998] MEYEROWITZ, E.M. (1998). *Genetic and molecular mechanisms of pattern formation in Arabidopsis flower development*. J. Plant Res., 111:233–242.
- [MÜNSTER et al. 1997] MÜNSTER, THOMAS, J. PAHNKE, A. DI ROSA, J. T. KIM, W. MARTIN, H. SAEDLER und G. THEISSEN (1997). *Floral Homeotic Genes Were Recruited From Homologous MADS-Box Genes Preexisting in the Common Ancestor of Ferns and Seed Plants*. Proc.Natl.Acad.Sci. USA, 94:2415–2420.
- [NEE und MAY 1997] NEE, SEAN und R. M. MAY (1997). *Extinction and the Loss of Evolutionary History*. Science, 278:692–694.

- [NEVERS et al. 1986] NEVERS, P., N. SHEPHERD und H. SAEDLER (1986). *Plant Transposable Elements*. Advances in Botanical Research, 12:103–203.
- [NYFFELER 1999] NYFFELER, RETO (1999). *A New Ordinal Classification of the Flowering Plants*. Trends Evol. Ecol., 14:168–170.
- [PARKINSON et al. 1999] PARKINSON, CHRISTOPHER L., K. L. ADAMS und J. D. PALMER (1999). *Multigene Analyses Identify the Three Earliest Lineages of Extant Flowering Plants*. Current Biology, 9:1485–1488.
- [PUTNAM 1993] PUTNAM, HILARY (1993). *Von einem realistischen Standpunkt*. Rowohlt.
- [QIU et al. 1999] QIU, YIN-LONG, F. LEE, JUNGHO AMD BERTANSONI-QUADRONI, D. E. SOLTIS, P. S. SOLTIS, M. ZANIS, E. A. ZIMMER, Z. CHEN, V. SAVOLAINEN und M. W. CHASE (1999). *The Earliest Angiosperms: Evidence from Mitochondrial, Plastid and Nuclear Genomes*. Nature, 402:404–407.
- [RAFF 1996] RAFF, R.A. (1996). *The shape of life - Genes, development, and the evolution of animal form*. The University of Chicago Press, Chicago, USA.
- [SAEDLER und THEISSEN 1994] SAEDLER, H. und G. THEISSEN (1994). 'On the origin of species': Mythologische und molekularbiologische Vorstellungen zur Evolution von Mais. In: *Leopoldina Jahrbuch 1993*, S. 261–275, Halle / Saale. Deutsche Akademie der Naturforscher Leopoldina.
- [SCHWARZ-SOMMER et al. 1990] SCHWARZ-SOMMER, Z., P. HUIJSER, W. NACKEN, H. SAEDLER und H. SOMMER (1990). *Genetic control of flower development by homeotic genes in Antirrhinum majus*. Science, 250:931–936.
- [SMITH et al. 1992] SMITH, J. STEPHEN C., O. S. SMITH, S. WRIGHT, S. J. WALL und M. WALTON (1992). *Diversity of U.S. Hybrid Maize Germplasm As Revealed by Restriction Fragment Length Polymorphisms*. Crop Science, 32:598–604.
- [SMITH und SMITH 1992] SMITH, O.S. und J. SMITH (1992). *Measurement of Genetic Diversity among Maize Hybrids: A Comparison of Isozymic, RFLP, Pedigree, and Heterosis Data*. Maydica, 37:53–60.
- [SOLOW et al. 1993] SOLOW, ANDREW R., S. POLASKY und J. BROADUS (1993). *On the Measurement of Biological Diversity*. Journal of Environmental Economics and Management, 24:60–68.
- [SOLTIS et al. 1999] SOLTIS, PAMELA S., D. E. SOLTIS und M. W. CHASE (1999). *Angiosperm Phylogeny Inferred from Multiple Genes as a Tool for Comparative Biology*. Nature, 402:402–404.
- [SOMMER et al. 1990] SOMMER, H., J.-P. BELTRÀN, P. HUIJSER, H. PAPE, W.-E. LÖNNIG, H. SAEDLER und Z. SCHWARZ-SOMMER (1990). *Deficiens, a homeotic gene involved in the control of flower morphogenesis in Antirrhinum majus: the protein shows homology to transcription factors*. EMBO J., 9:605–613.
- [SOULÉ 1991] SOULÉ, MICHAEL E. (1991). *Conservation: Tactics for a Constant Crisis*. Science, 253:744–750.
- [TAUTZ 1996] TAUTZ, D. (1996). *Selector genes, polymorphisms, and evolution*. Science, 271:160–161.
- [THEISSEN et al. 2000] THEISSEN, G., A. BECKER, A. DI ROSA, A. KANNO, J. KIM, T. MÜNSTER, K.-U. WINTER und H. SAEDLER (2000). *A short history of MADS-box genes in plants*.
- [THEISSEN et al. 1996] THEISSEN, GÜNTER, J. T. KIM und H. SAEDLER (1996). *Classification and Phylogeny of the MADS-Box Multigene Family Suggest Defined Roles of MADS-Box Gene Subfamilies in the Morphological Evolution of Eukaryotes*. J.Mol.Evol, 43:484–516.
- [THEISSEN und SAEDLER 1995] THEISSEN, GÜNTER und H. SAEDLER (1995). *MADS-Box Genes in Plant Ontogeny and Phylogeny: Haeckel's 'Biogenetic Law' Revisited*. Current Opinion in Genetics and Development, 5:628–639.
- [THEISSEN und SAEDLER 1998] THEISSEN, GÜNTER und H. SAEDLER (1998). *Molecular architects of plant body plans*. Progress in Botany, 59:227–256.

- [VOS et al. 1995] VOS, P., R. HOGERS, M. BLEEKER, M. REIJANS, T. VAN DE LEE, M. HORNES, A. FRIJTERS, J. POT, J. PELEMAN, M. KUIPER und M. ZABEAU (1995). *AFLP: A New Concept for DNA Fingerprinting*. Nucl. Acids Res., 23:4407–4414.
- [WANG et al. 1999] WANG, R.-L., A. STEC, J. HEY, L. LUKENS und J. DOEBLEY (1999). *The limits of selection during maize domestication*. Nature, 398:236–239.
- [WEIGEL und MEYEROWITZ 1994] WEIGEL, D. und E. MEYEROWITZ (1994). *The ABCs of floral homeotic genes*. Cell, 78:203–209.
- [WHITKUS et al. 1998] WHITKUS, R., M. DE LA CRUZ, L. MOTA-BRAVO und A. GÓMEZ-POMPA (1998). *Genetic Diversity and Relationships of Cacao (Theobroma cacao L.) in Southern Mexico*. Theor Appl Genet, 96:621–627.
- [WILLIAMS et al. 1991] WILLIAMS, PAUL H., C. J. HUMPHRIES und R. I. VANE-WRIGHT (1991). *Measuring Biodiversity Taxonomic Relatedness for Conservation Priorities*. Australian Systematic Botany, 4:665–680.
- [WILLSON 1998] WILLSON, STEPHEN J. (1998). *Measuring Inconsistency in Phylogenetic Trees*. J. theor. Biol, 190:15–36.
- [WILSON 1993] WILSON, EDWARD O. (1993). *Biodiversity: Challenge, Science, Opportunity*. Amer. Zool., 34:5–11.
- [WINTER et al. 1999] WINTER, KAI-UWE, A. BECKER, T. MÜNSTER, J. T. KIM, H. SAEDLER und G. THEISSEN (1999). *MADS-Box Genes Reveal That Gnetophytes Are More Closely Related to Conifers than to Flowering Plants*. Proc. Natl. Acad. Sci. USA, 96:7342–7347.
- [YANOFSKY et al. 1990] YANOFSKY, M.F., H. MA, J. BOWMAN, G. DREWS, K. FELDMAN und E. MEYEROWITZ (1990). *The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors*. Nature, 346:35–39.
- [YU und PAULS 1992] YU, KANGFU und K. PAULS (1992). *Optimization of the PCR Program for RAPD Analysis*. Nucl. Acids Res., 20:2606.
- [ZHU et al. 1998] ZHU, J., M. GALE, S. QUARRIE, M. JACKSON und G. BRYAN (1998). *AFLP Markers for the Study of Rice Biodiversity*. Theoretical and Applied Genetics, 96:602–611.