

Genome im Überblick

Laurence Loewe, Döbelestraße 34, 78462 Konstanz

Zusammenfassung: Die Entwicklung und Optimierung von Verfahren zum Sequenzieren großer Mengen der Erbsubstanz DNA ermöglichte die vollständige Bestimmung der Reihenfolge der Bausteine des Genoms, also der Gesamtheit aller Gene, von bisher fünf Mikroorganismen. Dazu gehören die kleinste bekannte Zelle überhaupt, *Mycoplasma genitalium*, die Eubakterien *Haemophilus influenzae* und *Escherichia coli*, das Archaeobakterium *Methanococcus jannaschii* sowie die eukaryote Bierhefe *Saccharomyces cerevisiae*. Die Sequenzierung der Genome weiterer Modellorganismen und des Menschen ist bereits in vollem Gange. Aus Sicht der Ursprungsforschung sind die Informationen in den Genomen und ihre jeweiligen Ähnlichkeiten von Interesse. Zum einen gibt es neben einigen anderen Besonderheiten erstaunliche Merkmalskonflikte bei den Archaeobakterien. Zum anderen läßt die einfachste bekannte Zelle, *Mycoplasma genitalium*, erkennen, daß selbst die einfachsten modernen Lebewesen scheinbar immer noch einige hundert Proteine benötigen. Wie bei allen untersuchten komplexen Modellorganismen festgestellt wurde, scheinen nicht alle identifizierten Gene für die Vermehrung unter Laborbedingungen absolut unerlässlich zu sein. Schätzungen deuten darauf hin, daß nur in etwa 30% aller Fälle die Abwesenheit eines Gens tödliche Folgen hat. Aus evolutionstheoretischer Perspektive bleibt momentan unklar, wie die Gene entstanden sein sollen, bei deren Fehlen ein Überleben des Lebewesens unmöglich wäre. Auch die Entstehung der Gene, welche nicht spürbar zu einer Erhöhung der durchschnittlichen Nachkommenzahl führen, ist unklar. Auf jeden Fall ist es an der Zeit, sich an eine „Biologie der Vogelperspektive“ zu gewöhnen: Genome im Überblick.

Einleitung

Mit der Aufklärung der räumlichen Struktur der Erbsubstanz DNA im Jahre 1953 und des genetischen Codes in den 60er Jahren sind wesentliche Grundsteine für die moderne Molekularbiologie gelegt worden. Einige wesentliche Erkenntnisse sind in Abb. 1 zusammengefaßt. Doch auch wenn nun prinzipiell klar ist, daß die Reihenfolge der vier Basen in der DNA die Reihenfolge (Sequenz) der 20 Aminosäuren in den Proteinen und damit ihre Struktur

und Funktion bestimmt, so ist damit immer noch nicht geklärt, welche Gene und Proteine auf welche Weise miteinander wechselwirken, um in ihrer jeweiligen Zelle zu einem sinnvollen Ganzen zu werden.

Um dies herauszufinden, wurde in den letzten Jahrzehnten folgendermaßen verfahren: Gegeben sei ein bestimmtes, interessantes vererbbares Phänomen (z.B. eine Erbkrankheit). Nun suche man unter viel Aufwand nach den Proteinen und den Genen, welche für diese Eigenschaft verantwortlich sind, sofern es sie denn gibt. Hat man diese gefunden, so untersuche man ihren strukturellen und funktionellen Kontext: wo im Genom liegt das Gen, wie wird es vererbt, welche Umstände führen zur Aktivierung des Gens, wie wird die Arbeitskopie des Gens (mRNA) bearbeitet, bevor sie in ein Protein übersetzt wird, wo kommt das Protein zum Einsatz, welche drei-dimensionale Struktur besitzt es, welche biochemischen Funktionen hat es, womit interagiert es, und nicht zuletzt, welche Mutationen führen dazu, daß dieses Protein seine Funktion verliert. Natürlich ist ebenfalls von Interesse, ob ähnliche Proteine bzw. Gene in anderen Organismen vorkommen und wenn, worin diese Ähnlichkeit besteht.

Zum ersten Mal ist es möglich, DNA-Sequenzen ganzer Genome zu vergleichen.

Dieser genzentrierte Ansatz war bisher praktisch nicht zu umgehen und sehr erfolgreich. Der größte Teil unseres bisherigen Wissens über die molekularen Grundlagen des Lebens wurde von vielen Tausenden von Forschern auf diese Weise gesammelt. Dennoch ist diese Arbeitsweise nicht nur mühsam, sondern auch grundsätzlich ungeeignet, um einen Überblick über Genome und ihre Evolution zu gewähren, da sich immer nur die (zufällig) entdeckten Gene auswerten lassen. Über DNA-Sequenzen zwischen den Genen, die genaue Reihenfolge der Gene und über Gene mit bislang unbekannter Bedeutung läßt sich so nichts herausfinden.

Wie FLEISCHMANN et al. (1995), GOFFEAU et al. (1996) und die von ihnen angegebene Literatur beschreiben, stehen heute ausgefeilte Strategien zur Verfügung, die sogenannte Genomprojekte ermöglichen, also das Sequenzieren ganzer Genome von Anfang bis Ende. Ist dies geschehen, dann sind einzelne, besonders interessante Gene mit relativ wenig Aufwand zu finden – vorausgesetzt, ein

Glossar

Kursiv gesetzt sind Artnamen.

Archaea ist der Name von einer der drei „Domänen“ des Lebens. Organismen dieses Typs zeichnen sich auf den ersten Blick durch Bakterienähnlichkeit aus, besitzen aber auch viele Sequenzmuster, die sonst eher für Eucarya charakteristisch sind. Singular: Archaeon.

Bacteria ist der Name der „Domäne“ des Lebens mit den „klassischen“ Bakterien. Zellen dieses Typs zeichnen sich durch das Fehlen eines Zellkerns und bestimmte Sequenzmuster aus.

DNA, die „Erbsubstanz“. Ein unvorstellbar langes Molekül, welches die genetische Information durch die spezifische Abfolge (=Sequenz) seiner 4 Bausteine A(Adenin), T(Thymin), C(Cytosin) und G(Guanin) codiert. Es stehen immer zwei sich paarende Stränge gegenüber (Doppelhelix), wobei immer nur A und T bzw. C und G zusammenpassen. Siehe auch Abb. 1.

Domänen des Lebens: Mit der Entdeckung der Archaea wurde die bis dahin höchste taxonomische Einheit „Reich“ zu schwach, um die beobachteten Sequenzunterschiede in Worte zu fassen. Daher wurden von WOESE die Domänen Archaea, Bacteria und Eucarya eingeführt. Siehe auch Abb. 2.

Escherichia coli ist der Name eines Bakteriums, welches im menschlichen Darm lebt. Abb. 4 zeigt einige Zellen dieser Art.

Eucarya ist der Name der „Domäne“ des Lebens mit den „höheren“ Zellen („Eukaryonten“). Organismen dieses Typs (z.B. Pflanzen, Pilze, Tiere) zeichnen sich durch eukaryonte Zellen aus, deren Hauptmerkmal ein erkennbarer Zellkern ist.

Gene sind Abschnitte auf der DNA, die für Proteine codieren. Von der DNA-Sequenz wird eine mRNA-Arbeitskopie angefertigt, die an den Ribosomen gemäß dem genetischen Code in ein Protein (Kette von Aminosäuren) übersetzt wird. Siehe auch Abb. 1.

Genom: Die Summe aller DNA-Moleküle und damit auch aller Chromosomen (die alle Gene enthalten).

Haemophilus influenzae Rd ist der Name eines Bakteriums, das als Parasit in Menschen lebt. Verschiedene Stämme dieses Bakteriums verursachen verschiedene Krankheiten. Der Zusatz „Rd“ kennzeichnet einen Stamm, der für Menschen nicht infektiös ist (und von dem in diesem Artikel die Rede ist).

Kilobasen: siehe Sequenz

Megabasen: siehe Sequenz

Methanococcus jannaschii ist der Artnamen einer einfachen Zelle aus der Tiefe des Pazifik, die vor allem bei hohen Temperaturen wächst und Methan synthetisiert.

Modellorganismen sind von Natur aus durch nichts vor anderen Organismen ausgezeichnet. Sie werden nur von manchen Menschen besonders intensiv erforscht, in der Hoffnung, die für diesen Organismus gewonnenen Erkenntnisse auf möglichst viele andere Organismen übertragen zu können.

mRNA ist der molekulare Bote, der die Information über den Aufbau der Proteine von der DNA zu den Ribosomen bringt. Abkürzung für messenger RNA. Vgl. RNA.

Mycoplasma genitalium ist der Artnamen eines Bakteriums, das als Parasit in Menschen lebt. Es gilt als die Zelle mit dem kleinsten Genom.

Prokaryonte Zellen („Prokaryonten“) besitzen weder Zellkern noch andere Zellorganel-

le (z.B. Mitochondrien, Chloroplasten...). Zu ihnen gehören die Bacteria und die Archaea.

Proteine sind lange Ketten von (20 verschiedenen) Aminosäuren, die sich auf eine spezifische Weise falten, um eine charakteristische Funktion auszuführen. Die Reihenfolge der Aminosäuren ist in der DNA festgelegt. Siehe auch Abb. 1.

RNA ist ein der DNA sehr ähnliches Makromolekül mit mehreren entscheidenden Rollen bei der Informationsverarbeitung in der Zelle: mRNA ist die Arbeitskopie des Gens, tRNA's tragen die dem genetischen Code entsprechenden Aminosäuren (vgl. Abb. 1) und rRNA's übernehmen verschiedene spezielle Aufgaben in den Ribosomen.

Saccharomyces cerevisiae ist der Artnamen der Bierhefe, ein Eukaryont (s. Abb. 3).

Sequenzen: Sequenzieren bedeutet für DNA: Die genau festgelegte Reihenfolge der vier verschiedenen Bausteine der DNA wird durch eine experimentelle Methode für einen vorliegenden DNA-Abschnitt bestimmt. Die Reihenfolge nennt man Sequenz; sie ist vergleichbar einer Abfolge von Buchstaben. Die Länge einer Sequenz wird in Basenpaaren (=bp) gemessen; Kilobasen = 1000 DNA-Bausteine Sequenzlänge, Megabasen = Millionen Bausteine hintereinander. Siehe auch Abb. 1 und 3. Kernidee der modernen Sequenzierungsmethoden ist der zufällige Kettenabbruch von DNA-Tochtersträngen; die daraus resultierenden Längenunterschiede werden in einem geeigneten Medium (Sequenzier-Gel) durch elektrische Ladungen aufgetrennt.

Stück der Sequenz ist bereits bekannt. Allgemein wird erwartet, daß diese „Dienstleistung“ der biologischen (und v.a. pharmazeutischen) Forschung ungeheuren Auftrieb verleihen wird. Mittlerweile werden die ersten Früchte dieser Genomprojekte sichtbar: Fünf vollständig sequenzierte Genome wurden bereits veröffentlicht (FLEISCHMANN et al. 1995; FRASER et al. 1995; BULT et al. 1996; GOFFEAU et al. 1996; BLATTNER 1997) und weitere sollen folgen. Hier sollen einige Ergebnisse und Trends für die Ursprungsforschung angesprochen werden.

Modellorganismen

Da ein Genomprojekt erheblichen Aufwand mit sich bringt, werden in erster Linie solche Organismen bearbeitet, von denen man sich besondere medizinisch oder biologisch relevante Einsichten erhofft:

Neben dem Menschen und manchen seiner Parasiten sind dies bestimmte Modellorganismen, über die schon viel herausgefunden wurde und für deren weitere Erforschung wirkungsvolle Methoden zur Verfügung stehen. Letztlich ist man natürlich daran interessiert, möglichst viel über den Menschen herauszufinden. Modellorganismen tragen zu biologischer Erkenntnis bei, indem sie mit möglichst wenig Aufwand Experimente ermöglichen, die bei Menschen ethisch nicht zu verantworten wären. Damit ist die Hoffnung verbunden, möglichst viele der gewonnenen Erkenntnisse auf den Menschen übertragen zu können.

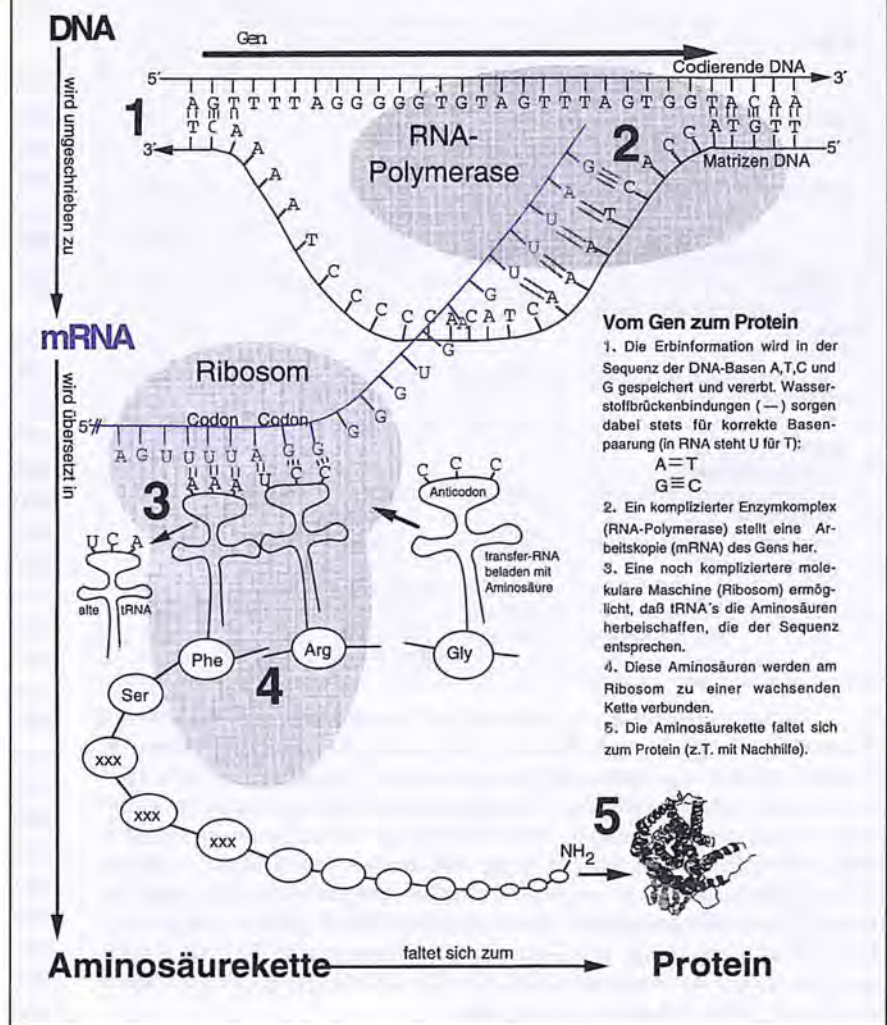
Bekannte Modellorganismen sind beispielsweise das Darmbakterium *Escherichia coli*, ein Vertreter der Prokaryonten; die Bierhefe *Saccharomyces cerevisiae*, ein Eukaryont; *Caenorhabditis elegans*, der durchsichtige, anatomisch einfache Fadenwurm, von dem man das Schicksal aller seiner ca. 1000 Körperzellen kennt (ALBERTS et al.

1994, S. 1067); *Drosophila melanogaster*, die Fruchtfliege; *Arabidopsis thaliana*, eine kleine krautige Pflanze, und *Mus musculus*, die Maus, der bisher realistischste, handhabbare Modellorganismus für den Menschen.

Die bisher veröffentlichten Genomsequenzen¹ repräsentieren neben dem kürzesten bekannten Genom (*Mycoplasma genitalium*) die drei sogenannten Domänen des Lebens nach WOESE (WOESE et al. 1990; BULT et al. 1996; MORELL 1996): *Haemophilus influenzae* und *Escherichia coli* (Bacteria), *Methanococcus jannaschii* (Archaea) und *Saccharomyces cerevisiae* (Eucarya). Eine schematische Darstellung der drei Domänen findet sich in Abb. 2.

Tab. 1 gibt einen Überblick über Systematik, Genomgröße und Zahl der Gene einiger aktueller Modellorganismen. Neben den bestehenden Unsicherheiten (vor allem in Bezug auf Anzahl und Wirkung der Gene) fällt auf, daß manche Organismen sehr stark schwankende Genzahlen besitzen können: Während Viren zwischen vier und mehr als hundert Gene besitzen können, benötigt die einfachste bekannte „selbständige“ Zelle, *Mycoplasma genitalium* bereits 470 Proteine.² Damit geben sich jedoch keinesfalls alle Bakterien (und Archaea) zufrieden: Während die meisten wohl zwischen 2000 und 4000 Gene haben, gibt es auch „Elefanten“ mit ca. 8000 Genen bei einer Genomgröße, die etwa 70% von der der Hefe beträgt. Eukaryontengenome scheinen stets mehr als 5000 Gene bei mehr als 10 Megabasen zu besitzen, Vielzeller allgemein mindestens mehr als doppelt so viele Gene bei mehr als 10facher DNA-Länge. Die Wirbeltiere, deren Genanzahl bisher abschätzbar war, scheinen mehr als 70.000 Gene zu tragen, wobei dieser Wert mit erheblichen Unsicherheiten behaftet ist. Trotz intensiver Erforschung insbesondere des menschlichen Genoms kann immer noch nicht sicher gesagt werden, ob der Mensch nun weniger als 50.000 oder mehr als 100.000 Gene trägt. In jedem Fall scheinen einige dieser Gene in Multigenfamilien vorzuliegen; es gibt beispielsweise ca. 1000 Phosphatasen und etwa 2000 Proteinkinasen im Menschen (MIKLOS et al. 1996). Zwei Proteine gehören dann zu einer Familie, wenn ihre Aminosäuresequenzen ein auffälliges Maß an Ähnlichkeit besitzen (z.B. mehr als ca. 30%). So gut wie immer kann daraus auch eine Ähnlichkeit in der Raumstruktur und oft auch in der Funktion abgeleitet werden (ORENGO et al. 1994). Aufgrund dieses Zusammenhangs konnte den vermuteten Proteincodierenden DNA-Sequenzen eines neu sequenzierten Genoms eine wahrscheinliche Funktion zugeordnet werden (vgl. Tab. 2 für *Mycoplasma genitalium* und FLEISCHMANN et al. (1995) bzw. BULT et al. (1996) für *Haemophilus influenzae* bzw. *Methanococcus jannaschii*).

Gene codieren Proteine



Woraus besteht ein Genom?

Ganz allgemein gesagt ist ein Genom zunächst einmal die Summe aller Gene eines Organismus und aller weiteren Sequenzen zwischen den Genen, die für Organisation und Regulation nötig sind. Physikalisch-chemisch gesehen besteht das Genom aus einem oder mehreren unheimlich langen (und daher unglaublich verwickelten) *ununterbrochenen* DNA-Fäden, die Chromosomen genannt werden. Prokaryontische Zellen besitzen meist ein einziges, ringförmiges Chromosom (neben wenigen kleinen DNA-Molekülen, die Plasmide genannt werden). Eukaryontengenome bestehen meist aus mehreren fadenartigen Chromosomen mit charakteristischen Organisationsmerkmalen an ihrem Zentrum und an ihren Enden. Wenn Eukaryontenzellen mehrere Kopien ihres Genoms enthalten, dann nennt man sie polyploid; meistens enthalten sie zwei Kopien von jedem Chromosom und sind somit diploid. Die Angaben über Genomlänge und Genanzahl in Tab. 1 beziehen sich auf eine einfache Kopie des Genoms (haploid).

Abb. 1: Aus der Information in den Genen werden Proteine gebaut. Wichtige Schritte auf diesem komplizierten Weg sind durch-numeriert. Die Darstellung ist nicht maßstabsgetreu.

	Genomlänge in Basenpaaren	Gene insg. (ca.)	Letale Gene (im Labor)	Gene in Protein- Familien
Viren				
Q β -Bakteriophage	4.220	4	$\approx 100\%$	$\approx 0\%$
T4-Bakteriophage	166.000	150		
Organismen				
Bacteria				
<i>Mycoplasma genitalium</i>	580.070	470		
<i>Haemophilus influenzae</i> Rd	1.830.137	1.743		$\approx 30\%$
<i>Bacillus subtilis</i>	4.200.000	3.700		
<i>Escherichia coli</i>	4.700.000	4.100		$\approx 46\%$
<i>Myxococcus xanthus</i>	9.450.000	8.000		
Archaea				
<i>Methanococcus jannaschii</i>	1.739.933	1.738		
Eucarya				
Pilze				
<i>Saccharomyces cerevisiae</i>	13.389.000	>5.885	$\approx 30\%$	$\approx 14\%$
Pflanzen				
<i>Arabidopsis thaliana</i>	≈ 100 Mio.	≈ 25.000	$? 2\%$	
Tiere / Mensch				
Fadenwürmer				
<i>Caenorhabditis elegans</i>	100 Mio.	14.000	$\approx 23\%$	
Gliederfüßer				
<i>Drosophila melanogaster</i>	165 Mio.	≈ 12.000	$\approx 30\%$	
Chordaten				
<i>Fugu rubripes</i>	400 Mio.	≈ 70.000	$?? 7\%$	
<i>Mus musculus</i>	3.300 Mio.	≈ 70.000	$? 25\%$	
<i>Homo sapiens</i>	3.300 Mio.	$? 50.000$		
		$? >100.000$		

Tab. 1: Systematik, Genomgröße (=Summe aller Chromosomen und Plasmide) und Genanzahl (nur vorhergesagte Proteine) einiger aktueller Modellorganismen im Überblick. Kürzlich sequenzierte Organismen sind kursiv gesetzt. Die Angaben entsprechen dem jeweiligen Stand der Forschung und sind auf ein idealisiertes Individuum der jeweiligen Art bezogen. Neue Erkenntnisse oder auch nur eng verwandte Individuen können daher zu geringfügig anderen Ergebnissen führen. „ $\approx$ “ deutet größere Ungenauigkeiten an, während „?“ auf die Fragwürdigkeit des besten bisher verfügbaren Wertes hindeutet. Quellenangaben: Alle Angaben aus MIKLOS et al. (1996) bis auf: Genomlänge und Genanzahl M.g.: FRASER et al. (1995), H.i.: FLEISCHMANN et al. (1995), M.j.: BULT et al. (1996), S.c.: GOFFEAU et al. (1996). Viren: Genomlänge: EIGEN (1993); Genanzahl: STRYER (1990).

Ein weiterer Gesichtspunkt, unter dem ein Genom betrachtet werden kann, ist die Art von Sequenzen, die es enthält. Dabei fallen vor allem drei große Gruppen auf (LEWIN 1994):

- Sequenzen, die nur einmal vorkommen (nicht-repetitiv). Hierzu gehören die meisten strukturellen Gene, auch wenn manche von ihnen Ähnlichkeiten zueinander besitzen und daher zu einer Proteinfamilie gehören.
- Sequenzen, die öfter vorkommen und länger sind (mäßig repetitiv). Diese Sequenzen sind oft durch Genduplikationen entstanden und können hinterher durch weitere Mutationen verändert worden sein, sie müssen also nicht völlig identisch sein. Hierzu gehören neben „gewöhnlichen“ Duplikationen beispielsweise auch Retrotransposons und Retroviren, die sich selbst im Genom vermehren.
- Sequenzen, die sehr oft vorkommen und kurz sind (hoch repetitiv). Manche dieser Sequenzen besitzen Funktionen an den Enden der Chromosomen und in ihrem Zentrum.

Drei der kürzlich sequenzierten Genome (die Bacteria und das Archaeon) besitzen auf ihrem ein-

zigen, ringförmigen Chromosom kaum repetitive Sequenzen, während diese bei Eukaryonten ausgesprochen häufig sein können. Dies schlägt sich auch in der Dichte der Gene auf der DNA nieder:

Während in den Prokaryonten etwas mehr als 1 Kilobase Sequenz pro Gen³ benötigt wird, benutzt die Hefe auf ihren 16 Chromosomen im Durchschnitt ca. 2 Kilobasen für ein Gen. Dennoch ist das Genom der Hefe sehr kompakt. *Caenorhabditis elegans* trägt ca. alle 6 Kilobasen ein Gen und beim Menschen sollen es gar 30 Kilobasen oder mehr sein (GOFFEAU et al. 1996).

Da ein Genom mindestens die Summe seiner Gene ist, wird in Tab. 2 ein Überblick über die 318 identifizierten Gene der einfachsten bekannten Zelle gegeben. Da für *Mycoplasma genitalium* 470 protein-codierende Regionen vorhergesagt wurden, müssen 152 Gene vorerst als unbekannt oder als nur hypothetisch zugeordnet gelten. Diese sind daher auch nicht in Tab. 2 aufgeführt, weshalb diese Tab. 32% kürzer ist, als sie eigentlich sein sollte – ein deutlicher Hinweis auf unser gegenwärtiges Nicht-Wissen. Hinter jeder größeren Funktions-Kategorie ist angegeben, wie viele Proteine von *Mycoplasma genitalium* dieser Kategorie zugeordnet werden konnten und wie viele es bei *Haemophilus influenzae* waren. Dies soll einen Vergleich mit Bakterien ermöglichen, die nicht ganz so von ihrer Umwelt abhängig sind wie das parasitäre *M. genitalium*. (Fast alle Aminosäuren und Glukose (oder evtl. auch andere Zucker) bezieht es von seinem Wirt.) Eine entsprechende Genliste für „normale“ Bakterien, die heutzutage auch alleine zurechtkommen, ist mehr als dreimal länger. Um einen Eindruck von der Fülle der vorhandenen und aufeinander abgestimmten Funktionen zu bekommen, genügt allerdings auch Tab. 2.

Besonderheiten der ersten fünf sequenzierten Organismen

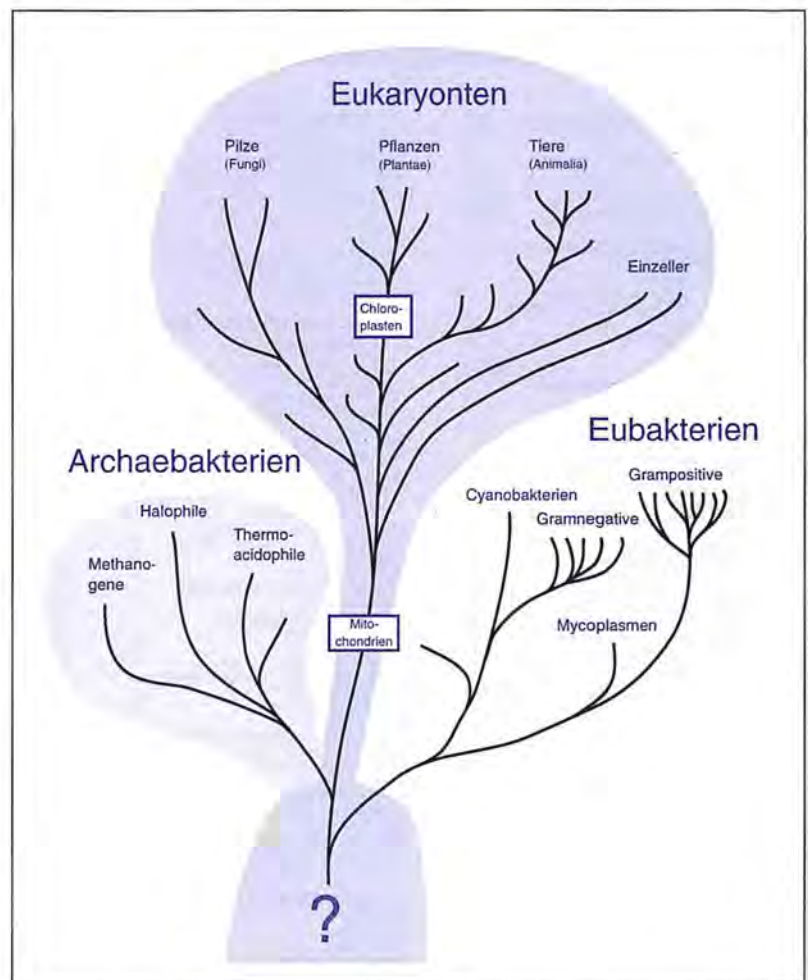
Haemophilus influenzae Rd ist der nicht-infektiöse Stamm von *Haemophilus influenzae*, einem nicht beweglichen, gramnegativen Bakterium, welches bisher nur im Menschen als natürlichem Wirt bekannt ist. Seine verschiedenen Stämme befallen vor allem die Atemwege (FLEISCHMANN et al. 1995). Obwohl im Vergleich zu *Mycoplasma genitalium* noch eine große Vielfalt an Stoffwechselwegen vorhanden ist, ist die parasitäre Lebensweise doch unverkennbar: Ein freilebender Organismus könnte sich wohl kaum auf Dauer den Verlust von drei Schlüsselenzymen des Zitratzyklus, der zentralen Drehscheibe des Stoffwechsels, leisten. *H. influenzae* erwartet statt dessen einfach 1g Glutamat pro Liter von seinem Wirt.

Mycoplasma genitalium ist ein sehr kleines, grampositives Bakterium mit dem kürzesten bekannten Genom und sehr speziellen Anpassungen an seinen Wirt. *M. genitalium* konnte seine Genomgröße außerordentlich stark reduzieren, da viele der benötigten Stoffwechselprodukte vom Wirt zur Verfügung gestellt werden. Dafür benötigt es aber für das Überleben unbedingt eine intelligente Abwehr des Wirts-Immunsystems, weshalb bis zu 4,5 % des Genoms für eben diesen Zweck verwendet werden könnten (GOFFEAU 1995). Da andere *Mycoplasma*-Arten (gleiche Gattung!) Genomgrößen mit mehr als 1 Megabase besitzen (DYBVIK et al. 1996), kann vermutet werden, daß die Reduktion der Genomgröße durch Mikroevolution erfolgte und bei den einzelnen Entwicklungslinien jeweils (dem Habitat entsprechend) zu unterschiedlichen Resultaten führte. Diese Hypothese sollte sich durch einen Vergleich der Anordnung der Gene im Genom überprüfen lassen: Durch Übereinanderlegen sollte sich eine ursprüngliche Gen-Anordnung festlegen lassen, die hauptsächlich durch unterschiedliche Deletionen (Verluste) zu den verschiedenen Genomen führt (andere Mutationen sollten dieses Bild in der gegebenen Zeit noch nicht allzu sehr verwischt haben können). Doch auch unabhängig von dieser Überlegung wird von einer sekundären Entwicklung des Parasitismus bei *M. genitalium* (und *H. influenzae*) ausgegangen (MANILOFF 1996). Das bedeutet, daß beide Organismen von Lebensformen abstammen (sollen?), die einmal weitgehend selbständig zurechtgekommen sind – ohne Wirt. Weiterhin bemerkenswert ist bei *Mycoplasma genitalium*:

Unerwartete Ordnung im Genom: Die Gene „rechts“ von dem vermuteten Replikationsursprung werden vorwiegend vom „plus“-Strang der DNA abgelesen, während die Gene „links“ davon meist vom gegenüberliegenden „minus“-Strang abgelesen werden (GOFFEAU 1995). Eine solche Ordnung ist in *Haemophilus influenzae* nicht zu beobachten (FRASER et al. 1995), wäre aber für weitere *Mycoplasma*-Arten durchaus zu erwarten, wenn diese wirklich durch Mikroevolution auseinander hervorgegangen sind.

Die Funktion vieler Gene ist unbekannt.

Abweichungen vom Universalen Genetischen Code: In *M. genitalium* codiert UGA für die Aminosäure Tryptophan, anstatt die Translation abubrechen (Stoppcodon). Bei einem aus makroevolutiver Sicht zu fordernden Übergang vom universalen zu diesem speziellen Code würden alle UGA-Stoppcodons nicht mehr das Ende eines Proteins bedeuten, was zumindest in einigen Fällen zu



Problemen bei der Faltung (und damit bei der Funktion) führen dürfte. ⁴

Methanococcus jannaschii, das Archaeon, wurde ursprünglich aus einem 2600 Meter tiefen Sediment in der Nähe eines „weißen Rauchers“ isoliert. Es wächst bei mehr als 200 bar und Temperaturen von 48–94 °C (Optimum bei 85 °C), verträgt keinen Sauerstoff und gewinnt seine Energie aus der Umsetzung von CO₂ mit H₂ zu Methan (BULT et al. 1996).

Abweichungen vom Universalen Genetischen Code: In *M. jannaschii* codiert UGA für die „21.“ Aminosäure Selenocystein, anstatt die Translation abubrechen (Stoppcodon).

Merkmalskonflikte zwischen Archaea, Bacteria und Eucarya: Ursprünglich wurden die Archaea als dritte Domäne des Lebens eingeführt, weil ihre ribosomale RNA eher den Eukaryonten glich als den damals bekannten Prokaryonten (Fox et al. 1977; WOESE et al. 1990). Diese Beobachtung hat sich bestätigt: Nicht nur die rRNA, sondern auch erstaunlich viele Proteine, die am Abschreiben, Übersetzen und Kopieren von DNA beteiligt sind, ähneln eher denen der Eucarya als denen der Bacteria – sofern es sich nicht gerade um archaeaspezifische Proteine handelt. In diesem Sinne hatte WOESE recht, als er die neue Domäne einführte. Im Energiestoffwechsel und in der Stickstofffixierung

Abb. 2: Schematische Darstellung der drei Domänen des Lebens. Nach TORTORA et al. (1992), S. 253.

Die vorläufig identifizierten Gene von *Mycoplasma genitalium*

Aminosäure-Biosynthese 1(68)		Pentose-Phosphat-Stoffwechsel		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%		%	
-------------------------------------	--	--------------------------------------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--	----------	--

traten dagegen auch Proteine zu Tage, die eher den Bacteria ähneln als den entsprechenden Vertretern der Eucarya (BULT et al. 1996). Nun stellt sich im makro-evolutiven Kontext die Frage, was ursprünglich und was konvergent evolviert ist: Stoffwechsel und grobe Zellstruktur oder Ribosomen und informationsverarbeitende Proteine? In beiden Fällen erscheint eine doppelte evolutive Entstehung sehr unwahrscheinlich, da viele unterschiedliche Proteine auf gleiche Weise hätten umgebaut werden müssen.⁵

Unbekannte Proteine: Nur 38% der vorhergesagten Gene von *Methanococcus* konnten sicher in den bekannten Datenbanken identifiziert werden. Damit haben mehr als die Hälfte aller Gene der Archaea eine bisher völlig unbekannte Funktion.

Saccharomyces cerevisiae, die Bierhefe, ist einer der einfachsten Eukaryonten. Einen Eindruck davon vermittelt Abb. 3. Die 16 Chromosomen dieses Pilzes lieferten schätzungsweise ca. 3000 Gene (50%), die keinem heute schon (woher auch immer) bekannten Protein zugeordnet werden können. Unter den wiedererkannten Genen lösten einige großes Erstaunen aus: Wer hätte vorher gedacht, daß fast die Hälfte der Proteine, von denen man weiß, daß ihre Defekte beim Menschen Erbkrankheiten auslösen, ein ähnliches Gegenstück in der Hefe besitzen? (Dies ist selbstverständlich für die Pharmaindustrie von besonderem Interesse...)

Die nächste große Herausforderung in der Hefeforschung wird sein, die Funktion aller 5885 vermuteten Gene im Detail aufzuklären. Eine immense Arbeit, die sich jedoch mit dem Vorliegen der Sequenz wesentlich einfacher gestaltet als bisher. Dabei ist eine der großen Fragen, welche biologische Bedeutung den vielen, ähnlichen Genen der Hefe in ihrem normalen Leben zukommt (GOFFEAU et al. 1996).

Escherichia coli, das menschliche Darmbakterium (Abb. 4) wurde erst als fünfter Organismus vollständig sequenziert. Obwohl an diesem Modellorganismus mehr Forscher arbeiteten als an irgendeinem anderen der oben genannten, dauerte die Fertigstellung wegen mangelnder Koordination erstaunlich lange. Hauptgrund dürfte wohl gewesen sein, daß die meisten Forscher *E. coli* nach dem eingangs genannten genzentrierten Ansatz untersucht haben. Daher ist aber auch schon von vielen *E. coli*-Genen aus dem Experiment bekannt, welche Funktion sie besitzen.

Auf der Suche nach der Urzelle

Die Kenntnis des vollständigen Genoms einer Zelle wirft auch noch auf ein anderes Gebiet besonderes Licht, vor allem, wenn dieses Genom von der einfachsten bekannten Zelle stammt: Wie entstand das Leben überhaupt? Seit langer Zeit wird diskutiert, welche minimalen molekularbiologischen Gege-



Abb. 3: *Saccharomyces cerevisiae*
(1600 fach)

benheiten vorliegen müssen, damit ein biologisches System überhaupt als „Leben“ bezeichnet werden kann. Dabei wird allgemein davon ausgegangen, daß diese Kernfrage beantwortet ist, sobald irgendein noch so primitives System vorliegt, welches seinen Phänotyp (Gestalt und Stoffwechsel) genetisch codieren kann und damit zu darwinischer Evolution befähigt ist (z.B. de DUVE 1994, S.235). Weitere Evolution nach DARWIN soll dann zu immer besser angepaßten Formen bzw. zum Erwerb neuer Strukturen führen. Viele Evolutionstheoretiker gehen davon aus, daß die Umweltbedingungen stets so geartet waren, daß sich die evolvierenden Lebewesen langsam den heutigen Formen nähern konnten. Da jedoch weder die jeweiligen Umweltbedingungen in der Erdgeschichte, noch die Reaktionen der damaligen Lebewesen darauf bekannt sind, liegt dieser vermutete Weg vom „Bakterium zum Menschen“ völlig im Dunkeln. Wie SCHERER (1996) gezeigt hat, zeichnen sich zudem erhebliche Probleme bei einer mechanistischen Erklärung dieser molekularen Evolution der Einzeller ab.

Mit dem Bekanntwerden des vollständigen Genoms von *Mycoplasma genitalium* wurden in diese Diskussion einige sehr wichtige Fakten eingebracht. Die Frage, wie viele Gene für die einfachste moderne Zelle nötig sind, hat heute eine einfache Antwort: höchstens 470, denn offensichtlich kommt *M. genitalium* damit aus – wenn auch nur in einem sehr speziellen Kontext, der für eine hypothetische Urzelle so nicht gegeben sein konnte, da *M. genitalium* Parasit ist, also „höheres“ Leben benötigt. MUSHEGIAN und KOONIN (1996) fragten sich nun, ob dieser Satz von Enzymen nicht weiter reduziert werden könnte, ohne das grundsätzlich moderne Design einer solchen hypothetischen Zelle zu verändern. Sie untersuchten die Genome von *Mycoplasma genitalium* und *Haemophilus influenzae* mit eigenen Methoden und identifizierten 468 bzw. 1703 Proteincodierende Gene. (Die Abweichungen zu vorher genannten Werten ergeben sich aus der Schwierig-



Abb. 4: *Escherichia coli*
(2000 fach)

keit, unbekannte Gene zu erkennen.) Nun nahmen sie an, daß *M. genitalium* als gram-positives und *H. influenzae* als gram-negatives Bakterium eine schon seit langem getrennte evolutionäre Geschichte haben, in der beide vor allem solche neuen Gene erwarben, die für die wirklichen Grundfunktionen der Zelle praktisch jederzeit ersatzlos gestrichen werden könnten. Daher wählten sie alle 240 Gene, die in beiden Bakterien gemeinsam vorkommen, als Basis für eine minimale, funktionierende, moderne Zelle aus. Nun nahmen sie noch zwei Korrekturen vor: Zum einen mußten 22 Proteine hinzugenommen werden, die zwar für die Stoffwechselwege von *M. genitalium* notwendig waren, aber keine Ähnlichkeit zu *H. influenzae* besaßen. Dafür wurden 6 Gene aus der resultierenden Liste gestrichen, weil sie sonst entweder doppelt vorkämen oder weil sie Funktionen ausübten, die für die parasitäre Lebensweise von *Mycoplasma genitalium* charakteristisch sind. Daraus ergibt sich eine Liste von 256 Proteinen, die für eine Zelle modernen Typs in etwa ausreichen soll. (Tab. 2 liefert einen Eindruck von der Menge und Vielfalt der 318 identifizierten Gene von *Mycoplasma genitalium*; die ursprünglich entdeckten (FRASER et al. 1995) Ähnlichkeiten zwischen *Mycoplasma genitalium* und *Haemophilus influenzae* sind mit „*“ markiert.)

Die einfachste bekannte Zelle besitzt immer noch mehrere hundert Gene.

Da die meisten dieser Proteine Ähnlichkeiten zu entsprechenden Vertretern bei den Archaea und bei den Eucarya besitzen, dies aber für sieben Schlüsselenzyme der DNA-Replikation nicht gilt, spekulieren MUSHEGIAN und KOONIN (1996), daß der letzte gemeinsame Vorfahr dieser drei Domänen ein Genom aus RNA hatte. Davon ausgehend schlugen sie vor, wie weitere Gene eingespart werden könnten. Falls die Ursuppe für eine ausreichende Konzentration von allen Aminosäuren, Fettsäuren, RNA- und DNA-Bausteinen sorgen könnte, und die betreffende Zelle sich mit einem RNA-Genom zufrieden geben könnte, würden möglicherweise auch 128 Gene ausreichen (PENNISI 1996). Einer so nahrhaften Ursuppe fehlt jedoch jede experimentelle und theoretische Grundlage, ebenso wie der Vermutung, die richtigen 128 DNA-Stücke und Proteine könnten dort je von selbst entstehen und zueinanderfinden (VOLLMERT 1985; SHAPIRO 1987; SHAPIRO 1988; JOYCE 1989).

Sollten diese Gene in ihrer Länge mit denen von *Mycoplasma genitalium* vergleichbar sein, dann ergäbe sich daraus eine minimale Genomlänge von ca. 160.000 Basenpaaren (bp). Dies ist eine außergewöhnliche Größe für ein RNA Genom, etwa 10 mal größer als die größten bekannten RNA-Viren und mit Sicherheit viel zu groß für realistische Ursuppenzenarien. Damit ein solches Genom auf Dauer

überlebensfähig wäre, benötigte es eine RNA-Polymerase, die mehr als 10 mal genauer kopieren würde als heute bekannte RNA-Polymerasen, und selbst dann würde diese Zelle sich ständig am Rande der Fehlerkatastrophe bewegen (EIGEN 1993). Daß diese eine ernste Bedrohung für Organismen ohne regelmäßige Rekombination darstellen kann, wurde nicht nur theoretisch (MAYNARD SMITH 1978; LYNCH et al. 1995), sondern auch experimentell nachgewiesen (ESCARMIS et al. 1996). Da jedoch vielfach betont wird (DE DUVE 1994), daß am Anfang der Evolution einfache und unspezifische (d.h. ungenaue) Enzyme standen, ist nicht einzusehen, wie eine solche Protozelle längere Zeiträume hätte überleben können. In diesem Zusammenhang mag von Interesse sein, daß manche makroevolutionäre Szenarien die steigende Kopiergenauigkeit der DNA-Replikation für den Anstieg der Anzahl der Gene entlang der Linie Prokaryont – Eukaryont – Wirbeltier verantwortlich machen (BIRD 1995).

Weitere Vorschläge und Experimente zur Bestimmung einer minimalen Genzusammenstellung führen zu längeren (und wohl auch funktionstüchtigeren) Genomen (MANILOFF 1996; PENNISI 1996), die jedoch alle noch unter der 1000-Gen-Marke liegen. Ob diese Minimalorganismen auch in freier Natur überleben könnten, fernab von allen künstlichen Nährmedien und natürlichen Wirten, steht auf einem völlig anderen Blatt, da einerseits das Genom von vielfach abhängigen Parasiten als Ausgangspunkt gewählt wurde und andererseits die durchgeführten Mutagenesestudien im Labor stattfanden (MANILOFF 1996). In beiden Fällen ist die Umgebung schonender als unter den instabilen Umständen des rauen Mikробenalltags. Schon *Mycoplasma genitalium* braucht trotz seiner 470 Gene praktisch alle 20 Aminosäuren und kann möglicherweise nur auf Glukose wachsen, da die Verarbeitungswege für geringfügig andere Zucker evtl. nicht vorhanden sind (FRASER et al. 1995). Anschaulich gesagt: Selbst in einem Wasserglas mit Haushaltszucker und typischer Multivitamin-tablette würde *Mycoplasma genitalium* verhungern (da dort z.B. die Aminosäuren fehlen).

Wieviele Gene braucht man mindestens, um...

... gerade zu überleben oder um auch noch flexibel zu sein? Nachdem oben diskutiert wurde, wieviele Gene für die (unter günstigsten Umständen) einfachste Zelle nötig sein könnten, soll hier kurz diskutiert werden, wie viele Gene für manche Organismen einer bestimmten Organisationshöhe unabdingbar sind. Hier sind keine gesicherten Angaben möglich, da man eigentlich jedes Gen einzeln und in bestimmten Kombinationen mit anderen ausschalten müßte. Zu den in Tab. 1 genannten groben Abschätzungen gelangt man jedoch durch Mutagenesestudien, in denen viele Gene der Reihe

nach ausgeschaltet wurden und bei denen man bei einem bestimmten Bruchteil beobachten konnte, daß dies zu einem Abbruch der Vermehrung bzw. der Embryonalentwicklung führte. In den Fällen, wo diese Angaben einigermaßen verläßlich erscheinen, ist etwa 1/3 aller Gene des jeweiligen Organismus für sein Überleben unter Laborbedingungen unerläßlich (MIKLOS et al. 1996). Die Evolution unabdingbarer Gene bringt nicht geringe Schwierigkeiten mit sich, da zum einen für den Beginn der Evolution überhaupt ein funktionsfähiges Set von Genen nötig ist, bei dem alle wirklich wichtigen Gene schon vorliegen müssen. Zum anderen können neue Gene nur evolvieren, wenn sie die Fitness der schon existierenden Zusammenstellung nicht gefährden. Dies bedeutet, daß sich die Unabdingbarkeit mancher Gene erst im Lauf der Zeit, zusammen mit anderen, an ihnen angreifenden Strukturen entwickelt haben mußte.⁶ In solchen Fällen müßten einige „glückliche Zufälle“ für das Zusammentreffen aller nötigen Mutationen sorgen. Ob es je Populationen gegeben hat, die groß genug waren, um alle diese Zufälle wahrscheinlich zu machen, kann in einigen Fällen bezweifelt werden (SCHERER 1996).

Wozu die übrigen 2/3 aller Gene von den Organismen auch über einen längeren Zeitraum im Genom behalten werden, ist eine der großen Fragen der Biologie der Genome (GOFFEAU et al. 1996; LANDER 1996). Laborbedingungen unterscheiden sich von natürlichen Bedingungen im wesentlichen durch ihre Kontrollierbarkeit (und damit meist Konstanz), die dazu führt, daß nur bestimmte Gene benötigt werden, während in der Natur unter gelegentlich auftretenden Extrembedingungen auch seltene Gene benötigt werden können. Solche Effekte lassen sich, genauso wie geringfügige Fitnessveränderungen, im Labor nur schwer nachweisen (MIKLOS et al. 1996).

Doch nicht alle Genverluste sind mindestens unter speziellen Umweltbedingungen letal. Manche Gene ergänzen sich in ihrer Wirkung, so daß ein Selektionsnachteil nur auftritt, wenn mehrere Gene ausfallen. Dies wird daran deutlich, daß die phänotypischen Konsequenzen bei Ausfall eines Gens stark davon abhängen, welche Gene das Individuum sonst noch trägt (MIKLOS et al. 1996). Solche (mehrfach vorhandenen und aktiven) Gene stehen unter entsprechend der Kopienanzahl verringertem effektivem Selektionsdruck; dies wirft folgende wichtige Frage auf:

Wie können Gene evolvieren, deren Verlust fast folgenlos ist?

Für das Prinzip der Selektion gibt es genau definierte Grenzen. Zum einen operiert Selektion auf mehreren Ebenen (Organismus, Zellen, selbstreplizierende Genomteile), die sich ins Gehege kom-

men können (SZATHMÁRY et al. 1995; HURST et al. 1996), zum anderen sind der Stärke der Selektion in realen biologischen Populationen enge Grenzen gesetzt (im Gegensatz zu weitverbreiteten Computersimulationen hierzu). Sobald die Selektion nämlich zu stark wird, stirbt die betroffene Population aus (BÜRGER et al. 1995), ein tragischer Prozeß, der sich heutzutage ungezählte Male beim Aussterben von Arten wiederholt.

Der einzige Beurteilungsmaßstab der Selektion ist die momentane Fitness, d.h. die Anzahl der gezeugten Nachkommen in der nächsten Generation. Alle Mutationen, die sich nicht in einem meßbaren Unterschied der Zahl der Nachkommen auswirken, sind daher neutral, selbst wenn sie in anderen Situationen später extrem nützlich oder schädlich sind. Neutrale Mutationen breiten sich also unabhängig von allen Selektionsprozessen aus (KIMURA 1985; KIMURA 1987), so daß auf diese Weise sowohl langfristig potentiell positive als auch langfristig potentiell negative Mutationen in einer Population fixiert werden können. Man nennt diesen Prozeß auch genetische Drift. Über diesen Prozeß können sich also Mutationen einschleichen, die nicht so nachteilig sind, daß sie sofort aus der Population beseitigt werden, aber auch nicht wirklich neutral (oder gar positiv), so daß sie wenigstens keinen Schaden anrichten. Solche geringfügig schädlichen Mutationen können auf Genome in kleinen Populationen langfristig katastrophal wirken, da sie sich praktisch mit der gleichen Geschwindigkeit ausbreiten wie neutrale Mutationen sich in ihrer Wirkung jedoch addieren (LYNCH et al. 1995).

Anschaulich gesagt: Ein Schmetterling, der schlechter sehen kann, als andere, hat zwar gelegentlich Nachteile im Leben, doch hindert ihn das nicht unbedingt am Zeugen von Nachkommen. Gleiches gilt, falls er schlechter fliegen kann. Wenn jedoch irgendwann einmal nur noch Individuen mit schlechten Augen und schlechten Flugfähigkeiten vorhanden sind (weil sich diese durch Gendrift ausgebreitet haben), dann hat die absolute Fitness dieser Art noch mehr abgenommen. Häufen sich solche Mutationen, dann stirbt die Art aus, da die Fähigkeit Nachkommen zu zeugen im Laufe der Zeit abnimmt bzw. verloren geht. Eine absolute Fitness von weniger als 1 kann sich keine Art auf Dauer leisten, da sonst nicht genügend Nachkommen gezeugt werden, um die Vorfahren zu ersetzen.

Gene, deren Totalausfall keine meßbare Wirkung auf die Anzahl der Nachkommen hat, stehen daher effektiv nicht unter Selektionsdruck. Ihre Evolution ist also von genetischer Drift geprägt und führt somit nicht zu einer funktionalen Anpassung der betreffenden Strukturen im Laufe der Zeit. Woher kommen solche Strukturen dann? Wenn nicht verborgene oder vergangene Selektionsdrücke dafür verantwortlich sind, bleibt diese Frage im Rahmen der bisherigen Makroevolutionshypothese offen. Welche evolutive Geschichte die einzelnen Gene