

PETER UETZ & THURE ETZOLD

Die EMBL-Reptiliendatenbank

Um Daten zur Reptiliensystematik an zentraler Stelle zu sammeln, haben wir die weltweit erste öffentlich zugängliche Internet-Datenbank mit Informationen zur Systematik und Verbreitung der Reptilien eingerichtet (<http://www.embl-heidelberg.de/~uetz/LivingReptiles.html>). Diese nicht-kommerzielle Datenbank ist als Forum gedacht, in dem Herpetologen und Terrarianer Informationen sammeln und anderen zur Verfügung stellen können. Die Bedeutung und Zukunft taxonomischer Datenbanken wird anhand von ähnlichen Datenbanken in der Molekularbiologie demonstriert.

Einleitung

Mehrere Millionen Tierarten bevölkern die Erde (MAY 1992). Niemand weiß es jedoch genau. Darunter sind ca. 8000 Reptilienarten, aber auch hier gibt es keine genauen Zahlen. Laufend werden neue Arten entdeckt, bekannte umbenannt oder in ihrem Rang (Arten oder Unterarten) Arten neu bewertet. Dies führt nicht zuletzt zu terminologischen Problemen: nach BAUER (1994) sind z.B. für 173 Geckoarten über 800 Namen vergeben worden (inklusive Schreib- und Grammatikfehlern)! Bei anderen Wirbeltieren sieht es nicht viel besser aus: beispielsweise stellen 45% aller Säugernamen Synonyme dar (MAY & NEE 1995).

Um diesem Notstand abzuhelpfen, schlagen wir die Einrichtung einer zentralen Reptilien-Datenbank vor, in der alle bekannten Arten mit bibliographischen Daten erfaßt werden sollten. Im Gegensatz zu gedruckten Katalogen kann eine solche Datenbank fortlaufend ergänzt, aktualisiert oder korrigiert werden. Um eine stets weltweite Verfügbarkeit zu gewährleisten, sollte eine derartige Datenbank außerdem ans Internet angeschlossen sein. Um die Machbarkeit und vor allem Brauchbarkeit einer solchen Datensammlung zu demonstrieren, haben wir eine Prototyp-Datenbank eingerichtet, die auf dem am EMBL entwickelten Sequence Retrieval System (SRS) beruht.

Vergleichbare Datenbanken entstehen in der systematischen Biologie nur zögerlich. Online-Datenbanken zur Wirbeltiersystematik gibt es unseres Wissens noch gar keine, wenn man von der Datenbank „Mammal Species of the World“ einmal absieht, die jedoch nur (?) die elektroni-

sche Version des gleichnamigen Buches von WILSON & REEDER (1993) darstellt. Die bislang größte Datenbank zur Wirbeltiersystematik mit ca. 12500 Arten, „Fishbase“, ist bisher nur auf CD-ROM erhältlich (FROESE & PAULY, 1995).

Das Paradigma molekularbiologischer Datenbanken

Die meisten Erfahrungen mit biologischen Datenbanken gibt es im Bereich der Molekularbiologie, in der seit Ende der achtziger Jahre Methoden zur raschen Bestimmung von DNA-Sequenzen entwickelt wurden. Bereits 1982 wurden fast gleichzeitig Sequenzdatenbanken in Europa (am EMBL) und in den Vereinigten Staaten gegründet. Aufgrund der dramatisch gestiegenen Datenmenge wurde die EMBL-Datenbank 1995 in ein eigens dafür gegründetes Tochterinstitut in Cambridge (England) ausgelagert, das European Bioinformatics Institute (EBI, <http://www.ebi.ac.uk>).

Diese Datenbanken sind heute so bedeutend, daß praktisch alle neu entdeckten Gene oder Proteine darin aufgenommen werden. Die meisten neuen Gensequenzen werden heute nicht einmal mehr auf Papier gedruckt. Bei deren Vorstellung in den entsprechenden Fachzeitschriften wird nur noch eine Datenbanknummer genannt, unter der die Sequenz abrufbar ist. In der Tat verlangen bereits viele molekularbiologische Zeitschriften für die Veröffentlichung neuer Gene eine solche „accession number“. Der Autor muß also die Sequenzdaten und möglichst eine Beschreibung des neuentdeckten Gens zu einer Datenbank schicken, bevor er solche Ergebnisse publiziert.

Entdeckt heute ein Molekularbiologe ein vermeintlich neues Gen (oder Protein), schaut er umgekehrt zuerst in einer der öffentlich zugänglichen Datenbanken nach, ob dieses Gen bereits bekannt ist. Zur Suche muß der Wissenschaftler nur ein kurzes Stück der DNA- bzw. Proteinsequenz oder andere relevante Informationen kennen und erfährt schon Sekunden oder Minuten nach deren Eingabe in den Computer, ob dieses oder ein verwandtes Gen bereits beschrieben wurde.

Herpetologische Literatur vs. Online-Information

Taxonomische Datenbanken stehen vor noch größeren Problemen als Sequenzdatenbanken, da sie Arten in ihrer ganzen Komplexität möglichst eindeutig beschreiben sollen. Trotzdem sind solche Datenbanken aus vielen Gründen dringend notwendig: die zoologische Literatur ist auf viel mehr Zeitschriften, Bücher und Schriftenreihen verteilt als die molekularbiologische Literatur, zumal sie auch eine viel längere Tradition hat. Neubeschreibungen werden oft in ausländischen Zeitschriften publiziert, die nur wenigen Spezialisten – oft nur nach wochenlangem Leihverkehr – zur Verfügung stehen. Überdies werden viele Beiträge von wissenschaftlichen Amateuren in kleinen nationalen Zeitschriften veröffentlicht, obwohl sie einen bedeutenden Teil zoologischer Erkenntnisse liefern. Um sich die geringe Verbreitung von Zeitschriftenpublikationen einmal zu vergegenwärtigen, schaue man sich nur einmal die DGHT-Mitglieder- und damit Salamandra-Abonnentenzahlen in den USA an – es sind genau 48.

Demgegenüber gibt es mittlerweile weit mehr Internetcomputer als herpetologische Bibliotheken. Die Zeitschriftendatenbank des Deutschen Bibliotheksinstituts nennt Bibliotheken in nur acht deutschen Städten, die beispielsweise das „Journal of Herpetology“ abonniert haben. Dagegen bieten heute alle Universitätsbibliotheken Internetzugänge an. Während die DGHT zwar über 6000 Mitglieder und somit Salamandra-Abonnenten hat, sind mittlerweile schon mehr als eine Million Deutsche Abonnenten von T-Online/BTX und dadurch potentielle Nutzer von Online-Datenbanken. Mit anderen Worten: die Wahrscheinlichkeit einen Internetanschluß in Ihrer Nähe zu haben, ist um ein vielfaches größer als für ein bestimmtes herpetologisches Printmedium.

Während Zeitschriften stets aktuelle Informationen anbieten können, fallen Bücher schon schnell der „Ent-Aktualisierung“ anheim. Ge-

druckte „Datenbanken“ wie die Reihe „Das Tierreich“ sind nach einigen taxonomischen Revisionen und damit meist nach einigen Jahren unvollständig oder gar veraltet.

Nicht zuletzt deshalb wurden durch diverse Organisationen und Institute immer wieder große Datenbankprojekte zur Biodiversität angekündigt. Bis heute hat sich daraus jedoch noch keine konkrete Datensammlung ergeben, die auch öffentlich zugänglich ist (MACILWAIN 1994). Aus diesen Gründen haben wir begonnen, eine Datenbank zur Reptiliensystematik aufzubauen, die zunächst die Namen, Synonyme und Verbreitungsgebiete aller Reptilienarten erfassen und diese in beliebigen Kombinationen abfragbar machen soll. Es sei betont, daß diese Datenbank herkömmliche Publikationen weder ersetzen will noch kann. Sie könnte aber eine sinnvolle Ergänzung zur zentralen Sammlung taxonomischer Informationen sein, die das Fundament jeder biologischen Disziplin ausmachen.

Die Datenbank zur Reptiliensystematik am EMBL

Das EMBL unterhält zusammen mit dem European Bioinformatics Institute in Cambridge eine ganze Reihe von molekularbiologischen Datenbanken. Wir haben deshalb für unsere Datenbank deren Grundstruktur übernommen und an die herpetologischen Bedürfnisse angepaßt. Dadurch werden zwar nicht alle Spezialbedürfnisse befriedigt, es ergibt sich aber der Vorteil, daß verschiedene Datenbanken miteinander verknüpft werden können. Als Beispiel kann man sich zu allen Arten der Gattung *Sceloporus* neben den Verbreitungsgebieten auch DNA-Daten anzeigen lassen (von 23 *Sceloporus*-Arten sind schon DNA-Sequenzen bekannt!). In der Datenbank sind derzeit ca. 7500 Reptilienarten verzeichnet, wobei zu ca. 75% Verbreitungsangaben vorliegen (Abb. 1).

Bei der in Abb. 1 gezeigten Suche nach amerikanischen *Tantilla*-Arten erhält man beispielsweise folgende Ausgabe:

REPTILIA: *Tantilla atriceps*
 REPTILIA: *Tantilla coronata*
 REPTILIA: *Tantilla gracilis*
 REPTILIA: *Tantilla hobartsmithi*
 REPTILIA: *Tantilla nigriceps*
 REPTILIA: *Tantilla oolitica*
 REPTILIA: *Tantilla planiceps*
 REPTILIA: *Tantilla relicta*
 REPTILIA: *Tantilla rubra*
 REPTILIA: *Tantilla wilcoxi*
 REPTILIA: *Tantilla yaquia*

Man kann sich diese Liste auch mit den jeweiligen Verbreitungsgebieten zusammen anzeigen

Abb. 1. Eingabemaske für Datenbankrecherchen. Ende Juni 1996 konnten die drei wichtigsten Datenfelder abgefragt werden (Name, Synonym, Verbreitung). In dem gezeigten Beispiel wird nach den amerikanischen Arten der Colubriden-Gattung *Tantilla* gesucht. Auf die anderen Funktionen wird hier nicht eingegangen, zumal sich diese auch ändern können.

lassen (was hier jedoch aus Platzgründen unterbleibt). Alternativ kann man auch einzelne Namen anklicken und erhält dann den entsprechenden Eintrag, beispielsweise

Species:

Tantilla coronata

Synonyms:

Distribution:

USA (SE-Louisiana, Mississippi, Alabama, Georgia, NW-Florida, South Carolina, North Carolina, Tennessee, W-Kentucky, CS-Virginia)

Comment:

References:

Conant (1978)

Wie man sieht, sind hier noch keine Synonyme eingetragen. Die Verbreitungsinformationen wurden in diesem Fall einer Karte in CONANT (1978) entnommen. Die genauen bibliographischen Angaben findet man in einer separaten Literaturliste.

Obwohl die Datenbank vom reinen Datenvolumen her noch relativ klein ist, können doch schon einige Fragen damit beantwortet werden, so nach dem Verbreitungsgebiet einer bestimmten Art oder nach den Arten eines bestimmten Landes. Daneben kann man herausfinden, zu welcher Familie eine bestimmte Art oder Gattung gehört.

Nachdem die der Datenbank vorausgegangenen WWW-Seiten bereits im Januar im Internet angekündigt wurden, haben sich die Benutzerzahlen stetig erhöht (Abb. 2). Die Mehrzahl der ca. 80 Anfragen pro Tag kommt jedoch aus dem nordamerikanischen Raum (Tab. 1).

Einladung zur Mitarbeit

Eine Datensammlung wie die hier anvisierte übersteigt natürlich die Kapazität einzelner Personen. Da unsere Datenbank vor allem von Amateurher-

petologen benutzt wird, wie aus dem in Tab. 1 gezeigten Benutzerprofil hervorgeht (in den USA vor allem Firmenangehörige), möchten wir deshalb auch interessierte Laien dazu einladen, Daten zur Datenbank beizusteuern.

Der Beitrag von freiwilligen Nicht-Herpetologen ist auch deshalb wesentlich, weil vor allem bereits publizierte und damit weitgehend akzeptierte Daten erfaßt werden sollen. Soweit das Urheberrecht respektiert wird, kann also der Laie selbst durch Abtippen von Verbreitungsangaben aus Reiseberichten oder dergleichen einen wichtigen Beitrag zum Aufbau der Datenbank leisten. Aus Gründen des Urheberrechts sollen bis auf weiteres keine Volltexte mit detaillierten Beschreibungen erfaßt werden.

Eingeladen sind besonders die Arbeitsgruppen der DGHT, sich an der Sammlung und Aufbereitung der Daten zu beteiligen. Die AGs könnten auch als Datenkoordinatoren für bestimmte Taxa oder geographische Areale agieren.

Falls Sie etwas zur Datenbank beitragen wollen, teilen Sie uns bitte Ihr Interessengebiet mit (Familie oder Gattung(en), geographische Region, aber keine einzelne Arten) und evtl. welche Literatur Ihnen zur Verfügung steht. Wir senden Ihnen dann eine Liste mit Arten oder Regionen, über die wir nur wenige Informationen haben (sofern Sie das nicht selbst auf unserer WWW-Seite nachschauen können). Zusätzlich erhalten Sie ein Informationsblatt für Mitarbeiter. Die Daten sollten Sie dann ergänzen und mit den entsprechenden Literaturbelegen versehen an uns zurückschicken. Wir nehmen diese Daten dann in die Datenbank auf und senden Ihnen den kompletten Satz an Einträgen zur nochmaligen Durchsicht zu.

Aus praktischen Gründen können wir bei Erstfragen keine größeren Auszüge aus der Daten-

bank ausdrucken und verschicken. Sie können diese aber direkt im Internet einsehen.

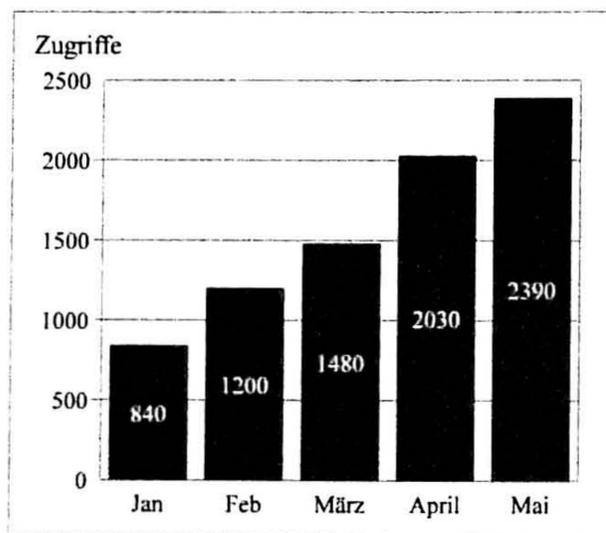


Abb. 2. Benutzerstatistik der EMBL-Datenbank im Zeitraum Januar bis Mai 1996. Die Datenbank wurde Anfang Januar in der newsgroup sci.bio.herp. angekündigt. Gezählt wurden die Zugriffe auf die Hauptseite (LivingReptiles.html). „Ein Zugriff“ entspricht gewöhnlich einer Sitzung am Computer.

Gründung einer AG „Reptiliendatenbank“?

Wir können angesichts der Größe der Aufgabe nur die Grundlagen legen und die technischen Mittel bereitstellen. Wie bei den molekularbiologischen Datenbanken hängt die Vollständigkeit und damit Brauchbarkeit einer systematischen Datenbank vor allem von den Beiträgen der Benutzer ab.

Bei genügendem Interesse wäre es sicher sinnvoll, langfristig eine AG „Datenbank“ (oder „Systematik“) zu gründen, deren Vorsitzender ein professioneller Taxonom sein sollte. Eine solche AG könnte sich auf das organisierte Sammeln von Daten konzentrieren. Deren Aktivitäten müßten natürlich mit ähnlichen (internationalen) Projekten koordiniert und verknüpft werden. Da 99% aller herpetologisch Interessierten Hobby-Herpetologen sind, ist die Einbeziehung dieser Amateure durchaus erstrebenswert. Auch durch kleine Beiträge vieler kann ein großes Projekt verwirklicht werden. Wir möchten hiermit die Diskussion über eine solche AG anregen und potentielle Leiter auffordern, weitere Argumente für oder gegen eine solche Gruppe vorzutragen.

Diskussion und Ausblick

Entwicklung und Einsatz von Datenbanken in der systematischen Biologie stehen erst am Anfang. In 10 oder 20 Jahren wird jedoch kein Zoologe mehr ohne Online-Datenbanken arbeiten können. Neben Texten werden auch Bilder und Töne abrufbar sein, in vielen Fällen ganze Filme. Im

	Herkunft	Zugriffe
com	Company	265
# #	unbekannt	254
net	Network	147
edu	Educational	123
org	Organisation	11
gov	Governmental	6
mil	Military	13
ca	Kanada	32
es	Spanien	30
it	Italien	18
us	USA	18
ch	Schweiz	17
no	Norwegen	15
se	Schweden	15
uk	England	15
de	Deutschland	10
ip	Japan	10
nl	Holland	9
fi	Finnland	6
au	Australien	5
fr	Frankreich	5
	weitere 17 Länder *	je 1-3

* Bulgarien, Belgien, Israel, Mexiko, Südafrika, Bahrain, Brasilien, Kroatien, Indonesien, Neuseeland, Portugal, Singapur, Slowenien, Tschechische Republik, Dänemark, Island, Malaysia

Tab. 1. Benutzerprofil der Reptiliendatenbank im Zeitraum 21.2.1996 - 10.3.1996 (nach Herkunft). Gezählt wurden 1037 Zugriffe auf die Hauptseite (LivingReptiles.html). Aus historischen Gründen wird das Internet in den USA in organisatorische Bereiche eingeteilt, so in Schulen und Universitäten („edu“) oder Regierungsbehörden („gov“). Der kommerzielle Bereich („com“) ist allerdings nicht mehr auf die USA beschränkt; mittlerweile benutzen auch viele nichtamerikanische Firmen diesen Domain-Namen. Knapp ein Viertel der Zugriffe (##) konnte nicht genauer zugeordnet werden.

Lauf der Zeit werden detaillierte Beschreibungen aller Arten mit aufgenommen, so daß man durch die Eingabe von einigen Schlüsselmerkmalen auch Tiere unbekannter Identität bestimmen kann. Theoretisch ist es sogar möglich, bei entsprechend aufbereiteten Artbeschreibungen automatisch Bestimmungsschlüssel zu generieren, in die man wiederum auch neue Arten leicht einfügen kann. Heute sind viele Bestimmungsschlüssel nach Beschreibung neuer Arten oft nicht mehr brauchbar. Datenbanken würden letztendlich auch bei Neubeschreibungen zur konsistenteren Benutzung von Merkmalen und Begriffen zwingen. Mit der Integration von Bildern ließen sich nicht nur viele ähnliche Arten der Unterarten direkt miteinander vergleichen und bestimmen, sondern auch die Untersuchung der Verbreitungsgebiete vereinfachen, indem man beispielsweise die Verbreitungskarten am Computer überlagert und so unmittelbar vergleichen kann. Das Hauptproblem hierbei ist nicht technischer Art, sondern der Aufwand, solche Karten laufend zu aktualisieren.

Verschiedene Datenbanken werden in der Zukunft miteinander verknüpft werden, wie morphologische mit molekularbiologischen oder ethologische mit neurobiologischen. Vielleicht werden in einer fernen Zukunft sogar morphologische Beschreibungen überflüssig sein, wenn die DNA-Sequenzen vieler Arten bekannt sind und man das Aussehen dieser Spezies direkt aus der DNA-Sequenz ableiten kann. In der Tat gibt es ja schon heute den Trend, phylogenetische Beziehungen durch Sequenzanalysen aufzuklären.

Die Integration kann auch auf nichtbiologische Datenbanken ausgedehnt werden, so daß als Beispiel zu einem bestimmten Verbreitungsgebiet zusätzlich Klimadaten abgefragt werden können. Das alles ist jedoch (z.T.) noch Zukunftsmusik. Bevor alle diese Möglichkeiten ausgeschöpft werden können, sollten die bereits herkömmlich publizierten Informationen elektronisch verfügbar gemacht werden. Hierzu ist nicht nur die Mitarbeit vieler Freiwilliger notwendig, sondern auch der (politische) Wille, solche Projekte zu fördern. Es ist keine Frage, ob herpetologische Datenbanken in der Zukunft erstellt werden. Es ist eher die Frage, ob wir uns daran beteiligen und diese mitgestalten oder ob wir dies anderen (beispielsweise Amerikanern) überlassen, von denen wir dann diese Datenbanken eventuell für teures Geld kaufen müssen.

Danksagung

Folgende Personen stellten ausführliche Informationen zur Verfügung: HERMANN SEUFER (Gek-

konidae), PETR NEČAS (Chamaeleonidae), KLAUS ADOLPHS (Cordylidae, Gerrhosauridae), WOLFGANG BISCHOFF (Lacertidae), WOLFGANG WÜSTER (Viperidae), JEAN JUST (Australische Reptilien).

Literatur

- BAUER, A.M. (1994): Gekkonidae (Reptilia, Sauria) Das Tierreich Bd. 109, Part I (Australia and Oceania), Walter de Gruyter, Berlin, 1-306.
- ETZOLD, T. & ARGOS, P. (1993): SRS – an indexing and retrieval tool for flat file data libraries. *Comput. Appl. Biosci.* 9:49-57.
- ETZOLD, T. & ARGOS, P. (1993): Transforming a set of biological flat file libraries to a fast access network. *Comput. Appl. Biosci.* 9: 59-64.
- FROESE, R. & PAULY, D. (1995): Fishbase: A Biological Database on Fish (CD-ROM + Manual), International Center for Living Aquatic Resources Management, Manila (<http://www.cgiar.org/iclarm/pubsof/pubsof.htm>).
- MACILWAIN, C. (1994): Global effort is launched to create taxonomic map of living organisms, *Nature* 368: 3.
- MAY, R. M. (1992): How many species inhabit the earth? *Sc. Am. Oct.* 1992: 18-24.
- MAY, R. M. & NEE, S. (1995): The species alias problem, *Nature* 378: 447-448.
- PATTERSON, B. D. (1996): The species alias problem, *Nature* 380: 589.
- WILSON, D. E. & D. M. REEDER (eds) (1993): *Mammal Species of the World*. Smithsonian Institution Press, 1206 pp. (<http://nmnhwww.si.edu/msw/>).

Dieser Artikel kann unter <http://www.embl-heidelberg.de/~uetz/articles/ReptilienDB.html> abgerufen werden.

Bei genügend großem Interesse kann die Datenbank auch als Diskettenversion verfügbar gemacht werden. Mitarbeiter (s.u.) erhielten die Datenbank dann verbilligt oder kostenlos (abhängig vom Umfang ihrer Beiträge).

Autoren

PETER UETZ
Europäisches Labor für Molekularbiologie (EMBL)
Meyerhofstr. 1, 69117 Heidelberg
E-Mail: uetz@embl-heidelberg.de/etzold@embl-heidelberg.de und
THURE ETZOLD
European Bioinformatics Institute, Hinxton Hall, GB - Cambridge CB 10 1RQ, E-Mail: s.o.