

Ulrich Schäfer und Herbert Jäckle,
Max-Planck-Institut für biophysikalische
Chemie, Abteilung für molekulare
Entwicklungsbiologie, Göttingen

Drosophila-Genomprojekte:

Vergangenheit, Gegenwart und Zukunft

Seit nahezu 100 Jahren haben genetische Studien mit *Drosophila melanogaster* wichtige Beiträge zum Verständnis biologischer Grundprozesse geleistet. Das Genom der Fruchtfliege ist unter den Eukaryoten eines der wenigen, dessen Kartierung und Sequenzierung abgeschlossen ist. Es zeigt sich, dass deutlich mehr als die Hälfte der bekannten menschlichen Gene, die in mutierter Form Krankheiten verursachen, im *Drosophila*-Genom konserviert sind. Für mehr als 80% der 13.601 annotierten *Drosophila*-Gene ist keine Funktion bekannt. Es werden Ansätze beschrieben, die zum Ziel haben, alle Gene von *Drosophila* funktional zu charakterisieren. Die Erkenntnisse werden eine wichtige Mittlerrolle für das Verständnis des humanen Genoms spielen.

► Die Fruchtfliege *Drosophila melanogaster* ist wahrscheinlich unter den höheren Eukaryoten derjenige Organismus, von dem wir – zumindest aus genetischer Sicht – die meiste Kenntnis haben. Das beruht natürlich zum großen Teil auf der über 90-jährigen Geschichte der Fruchtfliege und den daraus resultierenden unzähligen Beiträgen der so genannten „*Drosophila* community“ in Genetik, Cytologie, Biochemie, Neuro-, Entwicklungs- und Molekularbiologie, um nur einige Schwerpunkte zu nennen. Meilensteine im Hinblick auf das Genomprojekt waren die erste Genkarte überhaupt, die eine aus heutiger Sicht kaum erwähnenswerte Anzahl von sechs X-chromosomalen Loci enthielt, wie auch die erste physikalische Genkarte eines Organismus¹, die Polytänchromosomenkarte von Bridges (für eine kurze Übersicht hierzu, siehe Rubin und Lewis, 2000). Äußerst hilfreich, speziell für Neueinsteiger in die *Drosophila*-Genetik, war auch die gesammelte Beschreibung aller jemals produzierten Mutationen und chromosomalen Aberrationen, die seit ihrer ersten Zusammenfassung von Bridges und Brehme (1944) zwei Neufassungen in gedruckter Form erlebte, die letzte von Lindsley und Zimm im Jahre 1992. Seit 1993 steht im Rahmen von FlyBase (1999) eine stets aktualisierte, elektronische Version im Internet, die über entsprechende Links auch die direkte Bestellung einzelner Fliegenstämme ermöglicht (<http://flybase.bio.indiana.edu/>).

Die ersten *Drosophila*-Genomprojekte begannen Ende der achtziger Jahre zeitgleich in Europa und in den Vereinigten Staaten, wobei sich beide Seiten sinnvollerweise auf den euchromatischen Genom-Anteil beschränkten. Der heterochromatische Bereich, etwa ein Drittel des circa 180 Mb großen Genoms, besteht überwiegend aus mittel- bis hochrepetitiven Sequenzen und ist damit einer Klonierung und Kartierung kaum zugänglich. Obendrein enthält er bis auf wenige Ausnahmen keine Gene. 1989 wurden von einer amerikanischen Arbeitsgruppe und einem dafür gegründeten europäischen Konsortium erste physikalische Karten erstellt, die auf YAC-Klonen („yeast artificial chromosomes“ mit Insertionslängen um 150 kb; Garza *et al.*, 1989), beziehungsweise Cosmid-Klonen (Insertionslängen um 40 kb; Siden-Kiamos *et al.*, 1990) basierten. Letztere waren auch das Ausgangsmaterial für die Sequenzierung der telomernahen Region des X-Chromosoms durch das EDGP (European *Drosophila* Genome Project; Benos *et al.*, 2000). Zu Beginn der neunziger Jahre stieg dann das von G. M. Rubin geleitete BDGP (Berkeley *Drosophila* Genome Project) in die Genomanalyse ein. Eines der ersten Projekte war es, mit Hilfe von P1-Klonen (Insertionslänge um 80 kb) eine umfassende physikalische Karte zu erstellen, die etwa 90% des Genoms abdeckte (Kimmerly *et al.*, 1996). Klone hiervon wie auch aus den später produzierten BAC-Klonen („bacterial artificial chromosomes“ mit Insertionslängen von etwa 150 kb) mit autosomalen genomischen Fragmenten (Hoskins *et al.*, 2000) sollten anschließend vom BDGP zur Sequenzierung des Genoms verwandt werden.

1999 hat sich mit dem Einstieg von Celera Inc. die Sequenzierungsstrategie für das *Drosophila*-Genom drastisch geändert. In Abkehr von der ursprünglichen Strategie, also zuerst Kartieren und dann Sequenzieren von überlappenden Klonen, wollte Celera – als Testlauf für das menschliche Genom – das *Drosophila*-Genom nach dem bis dahin nur bei vergleichsweise kleinen, bakteriellen Genomen erfolgreiche „shotgun“-Verfahren vollständig sequenzieren. Dieser Versuch wurde zu Anfang mit sehr viel Skepsis betrachtet, hat dann aber, auch in Zusammenarbeit mit EDGP und besonders BDGP, letztendlich zum Erfolg geführt. Daher konnte die Sequenz des euchromatischen Anteils des *Drosophila*-Genoms im März 2000 publiziert werden (Adams *et al.*, 2000; Abb. 1). Wenn auch diese erste Genomsequenz noch nicht lücken- und fehlerfrei war, so war sie doch von einer solchen Genauigkeit und Vollständigkeit (deutlich weniger als 1 Fehler pro 10 kb in nichtrepetitiven Abschnitten), dass eine

Analyse der Genomstruktur und eine Genannotierung ohne weiteres möglich war (Rubin *et al.*, 2000b).

Welche Erkenntnisse über das *Drosophila*-Genom lassen sich aus der Gesamtsequenz gewinnen? Da ist zu allererst die Zahl von 13601 Protein-kodierenden Genen zu nennen. Jeweils knapp 40% dieser Gene sind durch cDNA-Klone (s. u.) und Ähnlichkeiten zu bekannten Proteinen anderer Spezies beziehungsweise nur durch nur eines von beiden abgesichert; der Rest beruht auf Computerprogrammen zur Vorhersage von Genen. Damit hat *Drosophila* nur gut doppelt so viele Gene wie die einzellige Hefe *Saccharomyces cerevisiae* (etwas über 6000 Gene) und deutlich weniger, als bei dem Fadenwurm *Caenorhabditis elegans* (circa 19000 Gene) gefunden wurden. Diese Zahlen sind auf den ersten Blick etwas überraschend, da die Fliege für eindeutig komplexer als der Wurm angesehen wird. Dabei ist auffällig, das Multidomänen-Proteine beim Wurm und bei der Fliege mit 2261 bzw. 2130 deutlich häufiger sind als bei der Hefe, die nur über 672 dieser komplexeren Proteine verfügt.

Die Fliege teilt etwa 16% ihrer Gene mit der Hefe. Viele dieser Gene kodieren für Proteine, die an grundlegenden zellulären Prozessen beteiligt sind, die allen Eukaryoten gemeinsam sind, wie etwa Zellteilung, Transkription und Translation. Mit *Caenorhabditis* hat *Drosophila* immerhin 35% der Gene gemein. Darunter sind auch solche, die für Komponenten von Signalkaskaden kodieren, wie sie typisch für multi-zelluläre Organismen sind. Im Vergleich zum humanen Genom liegen Daten nur für Gene vor, die in mutierter Form als Auslöser von Krankheiten beim Menschen identifiziert worden sind. Für die Untersuchung wurden 287 dieser Gene ausgewertet und 178 (62%) davon liegen konserviert auch in *Drosophila* vor (Fortini *et al.*, 2000). Kategorien mit höherer Repräsentation betrafen zum Beispiel Herzkrankheiten (6 von 6: 100%), Stoffwechselstörungen (14 von 17: 82%) und Krebs (47 von 65: 72%). Erwartungsgemäß sind Gene unterrepräsentiert sind, die an Störungen des endokrinen (12 von 31: 39%), des Immun- (7 von 17: 41%) oder des hämatologischen Systems (8 von 18: 44%) beteiligt sind, da diese Systeme in Fliege und Mensch nicht homolog sind. Das bedeutet, dass *Drosophila* auch für die Funktionsanalyse des menschlichen Genoms ein geeigneter „Modellorganismus“ sein kann (s. u.). Zudem steht die Fliege, zumindest molekular-evolutionär betrachtet, dem Menschen näher als der Wurm (Mushegian *et al.*, 1998).

Wie geht es nun weiter mit dem *Drosophila*-Genomprojekt? Eine erste Aufgabe wird sein, die Fehler und Lücken in der vor-

handenen Sequenz auszumerzen. Ein Schritt in diese Richtung war die kürzlich unter Federführung des BDGP veröffentlichte zweite Fassung der genomischen Sequenz. Parallel dazu werden von BDGP mehrere EST-Projekte (expressed sequence tags) und Sequenzierungen von Vollängen-cDNA-Klonen durchgeführt, um die nur durch Computervorhersagen postulierten Gene experimentell abzusichern oder zu widerlegen (Rubin *et al.*, 2000a). Bis März 2000 lagen mehr als 100.000 ESTs vor, die aus verschiedenen cDNA-Banken stammen. Erklärtes Ziel des BDGP ist es, noch einmal die doppelte Anzahl von ESTs zu erzeugen.

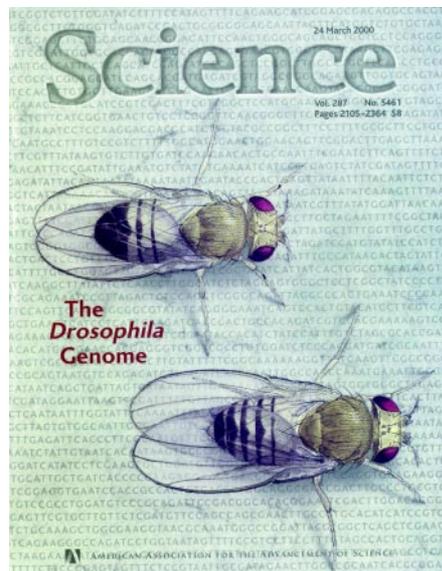


Abb. 1: Vorläufiger Höhepunkt des *Drosophila*-Genomprojektes war diese *Science*-Ausgabe vom 24. März 2000, die der Veröffentlichung der Genomsequenz sowie damit verwandten Artikeln gewidmet war.

Außer diesen mehr strukturellen Genomaspekten wird vor allem die funktionale Genomanalyse betont vorangetrieben. Denn selbst bei dem extrem gut untersuchten Modellorganismus *Drosophila* sind bis jetzt nur 10 bis 15% aller Gene molekular und funktional charakterisiert. Das heißt, für die meisten Gene ist es ungeklärt, welche Rolle sie spielen. Um diesem Mangel abzuwehren, wurden und werden in vielen Labors systematische Mutagenese-Ansätze durchgeführt. Neben der chemischen Mutagenese mit zumeist Ethylmethansulfonat wird hauptsächlich die Transposon-Mutagenese verwandt. Dabei beschränkt man sich nahezu ausschließlich auf solche, die sich von dem so genannten P-Element ableiten, einem natürlich vorkommenden *Drosophila*-Transposon. Per Keimbahnvermitteltem Gentransfer können die *in vitro*

veränderten P-Vektoren stabil in das Fliegen-genom integriert werden. P-Elemente inserieren bevorzugt in die Nähe des Transkriptionsstartes eines Gens (Spradling *et al.*, 1995). Das kann zu partiellen oder vollständigen Verlusten der Genaktivität und damit zu einer Mutation führen. Durch Aktivierung einer Transgen-abhängigen Transposase kann das integrierte P-Element remobilisiert werden. Kommt es zu einer Reversion des beobachteten Phänotyps, dann ist gewährleistet, dass die P-Insertion diesen Phänotyp verursacht hat.

Letale Einzelinsertionslinien für autosomale Gene werden schon seit längerem von BDGP erstellt und charakterisiert (Spradling *et al.*, 1995, 2000). Wir haben mit dem „Göttingen X-Chromosomen Projekt“ entsprechende Linien für X-chromosomale Gene erzeugt. Für mindestens jedes vierte essentielle Gen von *Drosophila* liegt derzeit eine P-Insertionsmutation vor. Da nur wenige Basenpaare der den P-Vektor flankierenden Sequenzen ausreichen, um die Integrationsstelle im vollständig sequenzierten Genom zu bestimmen, ist die Grundlage geschaffen für eine molekulare und funktionale Analyse des betroffenen Gens. Wegen der großen Ähnlichkeit zwischen den Genen von Fliege und Mensch können die so gewonnenen Daten auch erste Erkenntnisse über die Funktion des homologen menschlichen Gens liefern.

Nun gilt aber für die Mehrzahl der Transkriptionseinheiten, dass selbst ein Totalausfall nicht zu einem erkennbaren Phänotyp führt (Miklos und Rubin, 1996). Das mag daran liegen, dass diese Gene redundant vorliegen oder nur unter speziellen Bedingungen benötigt werden. Daher würden sie bei einer „normalen Mutagenese“ nicht entdeckt. Abhilfe könnte hier ein Ansatz schaffen, der auf Funktionsgewinn abzielt. Denn eine räumlich und/oder zeitlich vorgekommene falsche Expression eines Gens kann genetische Netzwerke derartig durcheinander bringen, dass diese negative Wechselwirkung phänotypisch zu erkennen ist. Erste Ansätze dieser Art sind bereits erfolgreich bei *Drosophila* durchgeführt worden (Rørth, 1996; Rørth *et al.*, 1998). Dabei ist das so genannte EP-Element (Rørth, 1996) eingesetzt worden, ein spezieller P-Vektor, der UAS-Stellen („upstream activating sequences“) enthält, die Bindungsstellen für den Hefe-Transkriptionsfaktor GAL4 darstellen. Diese UAS-Stellen sind im Vektor so orientiert, dass eine die Integrationsstelle flankierende Transkriptionseinheit aktiviert werden kann. Dieses geschieht durch ein weiteres Transgen, über das gewährleistet wird, dass GAL4 in der Fliege ektopisch exprimiert wird. In dieser Kombination führt dann eine EP-Insertion, die selbst keinen Phänotyp erzeugt hat,

durch Aktivierung oder Überexpression eines den UAS-Stellen nachgeschalteten Gens in 1 bis 2% der Fälle zur Letalität (Rørth *et al.*, 1998). Neuere Vektoren können diesen Wert um eine Größenordnung verbessern (eigene unveröffentlichte Beobachtungen). Mit Insertionslinien dieser Art stehen effektive Werkzeuge für eine funktionale Analyse der ansonsten phänotypisch „stillen“ Gene zur Verfügung.

Da es unmöglich ist, sämtliche *Drosophila*-Genomprojekte in dieser Übersicht angemessen zu berücksichtigen, möchten wir nur noch einige innovative Ansätze kurz erwähnen. Weitere Transposon-Mutagenesen bedienen sich anderer P-Vektoren mit speziellen Eigenschaften, sei es, um einzelne Gene zu charakterisieren (Lukacovich *et al.*, 2001), sei es, um das Genom systematisch mit 1 Mb großen, überlappenden Defizienzen abzudecken (Beumer *et al.*, 1998). Für die Gene, in die ein P-Element schwer oder gar nicht inseriert, bleibt dann noch die Mutagenese durch homologe Rekombination, eine bei Hefe und Maus gängige Technik zur gezielten Deletion von Genaktivitäten, die seit kurzem auch für *Drosophila* möglich ist (Rong und Golic, 2000). Etwas schneller mag es mit RNA-Interferenz-Ansätzen gehen (Kennerdell und Carthew, 1998); allerdings ist hierbei nicht immer klar, ob die beobachteten Phänotypen den Genverlust richtig widerspiegeln. Mehrere Gruppen versuchen auch, aufbauend auf der *Drosophila*-Sequenz und eigener Software Microarrays oder DNA-Chips herzustellen, um so eine Transkriptom-Analyse durchführen zu können. Darüber hinaus steht mittlerweile der erste kommerzielle *Drosophila*-Chip mit Oligonukleotiden, die alle 13601 Gene repräsentieren sollen, zur Verfügung (Affymetrix Inc.).

Die postgenomische Phase in der *Drosophila*-Forschung hat begonnen mit dem Ziel, letztlich die Funktion eines jeden Gens durch Mutantanalyse zu erkennen und die Befunde – aufgrund der konservierten Sequenzmotive – auf das menschliche Genom zu übertragen. Durch die Kombination chemischer Mutagenese mit konventionellen Rekombinationstechniken, Transposonmutagenese und Überexpressionsansätzen wird es möglich sein, innerhalb der nächsten Dekade Mutationen für jedes *Drosophila*-Gen zu erzeugen. Transkriptom-Analysen, kombiniert mit funktionellen Studien unter Einbeziehung der bei *Drosophila* üblichen „Modifier-Screens“ (z. B. Thomas und Wassarman, 1999) werden es erlauben, Querbeziehungen zwischen Genaktivitäten aufzudecken, um so genetische Regelkreise zu etablieren, die bestimmte Entwicklungsleistungen, Zellhomöostase, Zellerkennung und induktive Leistungen im Organismus gewährleisten. Für solche

Studien, die kurze Generationszeiten und einfache formalgenetische Handhabung des experimentellen Systems voraussetzen, um funktionelle Querbezüge zwischen Genaktivitäten rasch erkennen zu können, wird *Drosophila* als „Modellorganismus“ für funktionelle Beziehungen im Säuger genom und damit auch in Zukunft seiner Rolle als einer der „global player in life sciences“ verstärkt gerecht werden.

Literatur

- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D. et al.** (2000): The genome sequence of *Drosophila melanogaster*. *Science* 287: 2185-2195.
- Benos, P.V., Gatt, M.K., Ashburner, M., Murphy, L., Harris, D., et al.** (2000): From sequence to chromosome: the tip of the X chromosome of *D. melanogaster*. *Science* 287: 2220-2222.
- Beumer, K.J., Pimpinelli, S., and Golic, K.G.** (1998): Induced chromosomal exchange directs the segregation of recombinant chromatids in mitosis of *Drosophila*. *Genetics* 150: 173-188.
- Bridges, C.B. and Brehme, K.F.** (1944): The Mutants of *Drosophila melanogaster*. Carnegie Inst. Washington Publ. 552, Washington D.C.
- FlyBase** (1999): The FlyBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res.* 27: 85-88.
- Fortini, M.E., Skupski, M.P., Boguski, M.S., and Hariharan, I.K.** (2000): A survey of human disease gene counterparts in the *Drosophila* genome. *J. Cell Biol.* 150: F23-F29.
- Garza, D., Ajioka, J.W., Burke, D. T., and Hartl, D.L.** (1989): Mapping the *Drosophila* genome with yeast artificial chromosomes. *Science* 246: 641-646.
- Hoskins, R.A., Nelson, C.R., Berman, B.P., Laverty, T.R., George, R.A. et al.** (2000): A BAC-based physical map of the major autosomes of *Drosophila melanogaster*. *Science* 287: 2271-2274.
- Kennerdell, J.R., and Carthew, R.W.** (1998): Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell* 95: 1017-1026.
- Kimmerly, W., Stultz, K., Lewis, S., Lewis, K., Lustre, V., et al.** (1996): A P1-based physical map of the *Drosophila* euchromatic genome. *Genome Res.* 6: 414-430.
- Lindsley, D.L. and Zimm, G.G.** (1992): The Genome of *Drosophila melanogaster*. Academic Press, San Diego.
- Lukacovich, T., Asztalos, Z., Awano, W., Baba, K., Kondo, S., et al.** (2001): Dual-tagging gene trap of novel genes in *Drosophila melanogaster*. *Genetics* 157: 727-742.
- Miklos, G.L.G., and Rubin, G.M.** (1996): The role of the genome project in determining gene function: insights from model organisms. *Cell* 86: 521-529.
- Mushegian, A.R., Garey, J.R., Martin, J., and Liu, L.X.** (1998): Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes. *Genome Res.* 8: 590-598.

- Rong, Y.S., and Golic, K.G.** (2000): Gene targeting by homologous recombination in *Drosophila*. *Science* 288: 2013-2018.
- Rørth, P.** (1996): A modular misexpression screen in *Drosophila* detecting tissue-specific phenotypes. *Proc. Natl. Acad. Sci. USA* 93:12418-12422
- Rørth, P., Szabo, K., Bailey, A., Laverty, T., Rehm, J., et al.** (1998): Systematic gain-of-function genetics in *Drosophila*. *Development* 125: 1049-1057.
- Rubin, G.M., Hong, L., Brokstein, P., Evans-Holm, M., Frise, E., Stapleton, M., and Harvey, D.A.** (2000a): A *Drosophila* complementary DNA resource. *Science* 287: 2222-2224.
- Rubin, G.M., and Lewis, E.B.** (2000): A brief history of *Drosophila*'s contributions to genome research. *Science* 287: 2216-2218.
- Rubin, G.M., Yandell, M.D., Wortman, J.R., Miklos, G.L.G., Nelson, C.R. et al.** (2000b): Comparative genomics of the eukaryotes. *Science* 287: 2204-2215.
- Sidén-Kiamos, I., Saunders, R.D.C., Spanos, L., Majerus, T., Treanear, J., et al.** (1990): Towards a physical map of the *Drosophila melanogaster* genome: Mapping of cosmid clones within defined genomic divisions. *Nucleic Acids Res.* 18: 6261-6270.
- Spradling, A.C., Stern, D., Kiss, I., Roote, J., Laverty, T., and Rubin, G.M.** (1995): Gene disruptions using P transposable elements: An integral component of the *Drosophila* genome project. *Proc. Natl. Acad. Sci. USA* 92: 10824-10830.
- Spradling, A.C., Stern, D., Beaton, A., Rhem, E.J., Laverty, T., et al.** (1999): The BDGP Gene disruption project: single P element insertions mutating 25% of vital *Drosophila* genes. *Genetics* 153: 135-177.
- Thomas, B.I., and Wassarman, D.A.** (1999): A fly's eye view of biology. *Trends Genet.* 15: 184-190.

Korrespondenzadresse

PD Dr. Ulrich Schäfer

Max-Planck-Institut für biophysikalische Chemie
Abt. für molekulare Entwicklungsbiologie
D-37070 Göttingen
Tel.: 0551-201 1798
Fax: 0551-201 1755
eMail: uschaef@gwdg.de
<http://www.mpibpc.gwdg.de/abteilungen/170/>