

Erbgut verschiedener Nadelbäume vollständig entschlüsselt

Matthias Fladung, Karl Gebhardt, Birgit Kersten

Auf den ersten Blick erschien die Mitteilung, die am 22. Mai 2013 veröffentlicht wurde, nichts Aufregendes zu beinhalten: der Wissenschaftswelt wurde die Veröffentlichung des kompletten Erbguts (Genom) einer weiteren Pflanzenart bekanntgegeben [1]. Nach der Entschlüsselung des Genoms der Ackerschmalwand (*Arabidopsis thaliana*) im Jahr 2002 als erste Pflanze überhaupt, sind in den letzten zwölf Jahren unter anderem, auch Dank immer schnellerer und kostengünstigerer Methoden zur Genomsequenzierung, mehrere Duzend Pflanzengenome vollständig sequenziert worden. Doch der Fachmann erkannte die Bedeutung der Nachricht sofort: es handelt sich um das Genom der Gemeinen Fichte (*Picea abies*), das mit etwa 20 Gb (20×10^9 oder 20 000 000 000 Basenpaaren) etwa 7- bis 10-mal größer als das des Menschen und damit das größte Genom überhaupt ist, das bis heute sequenziert wurde.

Ein internationales Konsortium unter Beteiligung von Wissenschaftlern aus Schweden, Belgien und Nordamerika hat es in etwa drei Jahren geschafft, sowohl die Abfolge der Buchstaben des gesamten Erbguts der Fichte zu bestimmen als auch viele Gene, also die „aktiven“ Bestandteile, zu identifizieren [3].

Zum Vergleich: das erste Genom eines Baumes, das mit der Balsampappel (*Populus trichocarpa*) 2006 veröffentlicht wurde [4, 5], war mit etwa 520 Millionen Basenpaaren, verteilt auf 19 Chromosomen, vergleichsweise klein. Allgemein weisen Nadelbäume (Koniferen) mit Genomgrößen von knapp 10 bis zu 25 Gb die größten Genome im Reich der Pflanzen auf [6]. Im Vergleich dazu sind die Genome unserer wichtigsten Nahrungspflanzen maximal nur etwa ein Viertel so groß [7] (Abb. 2).

Warum aber wurde das Genom der Fichte für die Sequenzierung ausgewählt?

Koniferen, zu denen auch die Fichte zählt, stellen evolutionsbiologisch eine äußerst interessante Gruppe dar, da sie phylogenetisch sehr alt sind. Die phylogenetisch ältesten Pflanzen, die bisher sequenziert wurden, sind Pilze und Algen, die aber bekanntermaßen zu den niederen Pflanzen zählen. Fichten gehören trotz ihres hohen phylogenetischen Alters den höheren Pflanzen an und bisher ist noch kein Genom einer phylogenetisch so alten höheren Pflanze sequenziert worden.

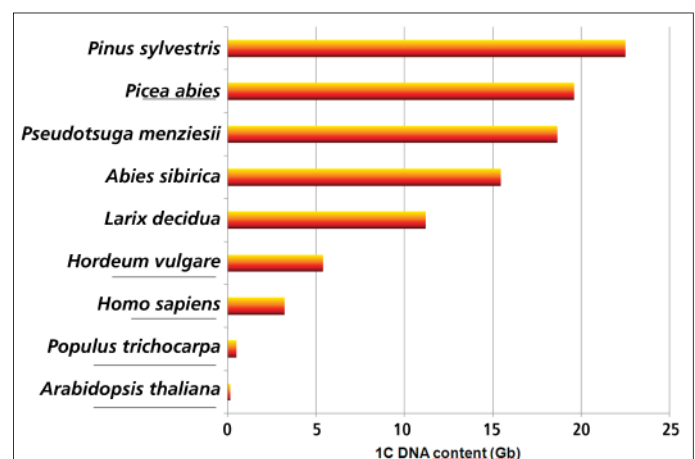


Abb. 1: Adulte 100-jährige Fichten (*Picea abies*)

In vielen terrestrischen Ökosystemen stellt die Fichte seit 200 Millionen Jahren die dominierende Pflanzenart dar, wie z. B. in borealen Wäldern, die einen 15 Mio km² großen Gürtel auf der Nordhalbkugel umspannen. Ökologisch bieten Fichtenwälder vielen Vogel- und Insektenarten, aber auch seltenen Pflanzenarten einen idealen Lebensraum.

Die Fortpflanzungsbiologie sowie die Genetik der Fichte sind in vielen Aspek-

Abb. 2: Genomgrößen ausgewählter diploider Pflanzen und des Menschen. Die Größe einer Kopie (1C) des jeweiligen Genoms entsprechend DOGS (<http://www.cbs.dtu.dk/databases/DOGS/>). Unterstrichene Speziesnamen bei Spezies mit (fast) vollständig publizierter Genomsequenz



M. Fladung und B. Kersten sind Mitarbeiter des Thünen-Institut für Forstgenetik, Großhansdorf. K. Gebhardt ist Mitarbeiter der Nordwestdeutschen Forstlichen Versuchsanstalt, Hann. Münden.



Matthias Fladung
matthias.fladung@ti.bund.de



Abb. 3: Weibliche Zapfen der gemeinen Fichte (*Picea abies*)



Abb. 4: Männliche Zapfen der gemeinen Fichte (*Picea abies*)

ten noch unverstanden. Wie ist es z. B. zu erklären, dass eine Pflanzenart scheinbar über Millionen von Jahren in ihrem Aussehen nahezu unverändert bleibt, während viele andere Arten phänotypisch einem starken Wandel unterworfen waren? Welche Gene steuern die Entwicklung vom Keimling zum adulten Individuum und welche die Differenzierung weiblicher und männlicher Zapfen (Abb. 3 und 4)? Aber auch wirtschaftliche und ökonomische Betrachtungen weisen die Fichte als eine der bedeutendsten Baumarten Europas aus. Fichtenholz findet nicht nur in der Möbelindustrie und im Gebäudebau Anwendung, sondern wird auch gerne für den Bau von Musikinstrumenten genutzt.

Warum ist das Genom der Fichte so groß?

Viele Pflanzenarten haben ihr Genom im Laufe der Evolution durch einfache Verdopplung des vorhandenen Chromosomensatzes vergrößert. Dadurch wird eine höhere „genetische Flexibilität“ erlangt, da nun von jedem Gen zwei Kopien vorliegen. Diese im Verlauf der Evolution erreichte „genetische Spielwiese“ erlaubt Mutationen in einer Kopie, die aber nicht zwangsläufig zum Tod der Pflanze führen, da die zweite Kopie gewissermaßen noch als Sicherheitskopie vorliegt.

Die besondere Größe des Genoms der Koniferen resultiert aber nicht aus solchen Genomverdopplungen, sondern aus einer während der Evolution langsam vorschreitenden Akkumulation von „sich wiederholender“ (repetitiver) DNA unbekannter Funktion und so genannten Retrotransposons („springenden Genen“), die zusammen einen Prozentsatz von 95 bis 98 % des Gesamtgenoms der Koniferen ausmachen.

Nur 3 % des Genoms besteht aus Protein-kodierenden Genen, also den Komponenten der DNA, die für die Merkmalsausprägung der Pflanzen verantwortlich sind. Die Analyse der Funktion dieser Gene wird aber durch das Vorliegen vieler inaktiver Gene („Pseudogene“), durch das Vorhandensein besonders langer, bis zu 10 000 Nukleotide umfassender Introns in den Genen sowie durch das Vorliegen von sehr großen Genfamilien erschwert.

Auch wenn alle aufgezählten Aspekte die Aufklärung der Funktion der Gene in der Fichte erschweren, erhofft man sich mit dem Vorliegen der Genomsequenz dennoch einen gewaltigen Schub für die Forstpflanzenzüchtung allgemein und für Koniferen insbesondere.

Die Anzahl der vorhergesagten Gene beträgt bei der Pappel 41 335 [8] und bei *Eucalyptus grandis* 36 376 [9], während bei der krautigen Pflanze *Arabidopsis thaliana* nur 27 416 Gene [10] identifiziert wurden. Trotz des sehr großen Genoms wurden für die Fichte erstaunlicherweise nur 28 354 Gene identifiziert [3], was vergleichbar mit der Anzahl der Gene des etwa 100-mal kleineren Genoms von *Arabidopsis* ist.

Bäume lassen sich nur schwer züchten

Bäume allgemein lassen sich im Vergleich zu krautigen Nutzpflanzen nur schwer züchten, da sie langlebig sind und häufig lange Reproduktionszyklen aufweisen. Mit der Veröffentlichung der Sequenz des ersten Baumgenoms, der Balsampappel (*Populus trichocarpa*) in 2006, erhoffte man sich, Gene zu finden, die an der Ausprägung Baum-spezifischer Eigenschaften wie Langlebigkeit, Angepasstheit an sich ändernde Umweltbedingungen und Holzbildung beteiligt sind [4]. Inzwischen

wurde tatsächlich eine Vielzahl „Baum-spezifischer“ Gene identifiziert, die jedoch nicht dem Bereich der Protein-kodierenden Gene zuzuordnen sind, sondern deren Genprodukte so genannte microRNAs (miRNAs) sind, die aber die Expression anderer Gene beeinflussen oder sogar unterdrücken und damit die Merkmalsausbildung beeinflussen können [11]. Andere Beispiele für regulatorische RNA-Moleküle sind so genannte kleine RNA-Moleküle (sRNA).

24 Nukleotide-lange sRNAs spielen eine entscheidende Rolle in der DNA-Methylierung von Genen, einem regulatorischen Prozess, welcher der Epigenetik zugeordnet wird. Für die Fichte konnte gewebespezifisch im Vergleich zu bedecksamigen Pflanzen keine oder eine nur vergleichsweise geringe Expression dieser kurzen sRNA-Moleküle gefunden werden, dagegen wurde eine Vielzahl von unbekanntem langen, aber nicht-kodierenden sowie extrem kurzen RNA-Molekülen (aus 21 Nukleotiden bestehend) entdeckt [3, 12]. Einige dieser RNAs spielen eine wichtige Rolle in der Abwehr von Pathogenen [13], andere prägen ein temperaturabhängiges Gedächtnis, das den Zeitpunkt des Austriebes und Triebabschlusses bei Fichten reguliert [15]. Auch die oben erwähnten Retrotransposons, die ursprünglich aus Viren stammen, könnten auf verschiedene Weise regulatorisch auf das Fichtengenom wirken [14].

Neben dem Genom des Zellkerns verfügen Pflanzen mit den Chloroplasten und den Mitochondrien in zwei weiteren Organellen über z. T. beträchtliche Mengen an DNA. Während die Chloroplasten den Apparat für die pflanzliche Fotosynthese beherbergen, stellen die Mitochondrien sozusagen die „Kraftwerke der Zelle“ dar. Die Sequenzierung des Fichtengenoms hat „nebenbei“ auch die Genomsequenz

dieser beiden Organellen offenbart: die Chloroplasten sind etwa 120 000 Basenpaare (bp) und die Mitochondrien etwa 4,8 Mb ($4,8 \cdot 10^6$ bp) groß [3]. Zusätzlich zur Fichtensequenz hat das schwedisch-amerikanisch-belgische Konsortium auch für andere Nadelbäume Genomsequenzen vorgestellt, allerdings in niedrigerer Auflösung und Genauigkeit. Dazu zählen die Gemeine Kiefer oder Waldkiefer (*Pinus sylvestris*), Sibirische Tanne (*Abies sibirica*), Europäische Eibe (*Taxus baccata*) und der Gemeine Wacholder (*Juniperus communis*).

Weitergehende Forschung

Aktivitäten anderer Arbeitsgruppen haben bereits ebenfalls zu vorläufig publizierten Genomsequenzen von Nadelbäumen geführt. Vor kurzem hat eine amerikanische Arbeitsgruppe die Entschlüsselung des Genoms der Weißfichte (*Picea glauca*) bekanntgegeben [15], eine andere amerikanische Arbeitsgruppe arbeitet intensiv an der Sequenzierung der Weihrauchkiefer (*Pinus taeda*) [16]. Es wird nur eine Frage der Zeit sein, wann Genomsequenzen anderer Nadelbaumarten wie Douglasie (*Pseudotsuga menziesii*) und Japanische Lärche (*Larix kaempferi*) oder Laubbaumarten wie Stieleiche (*Quercus robur*) und Amerikanische Buche (*Fagus grandifolia*) folgen.

Durch den Vergleich der Genome verschiedener krautiger und verholzender Pflanzen erhofft man sich sowohl einen immer tieferen Einblick in die Genomstruktur als auch Informationen über die Realisierung der Genominformation. Wichtige Merkmale wie Pflanzenwachstum, Produktivität, Qualität des Holzes, Gesundheit der Pflanze werden genetisch gesteuert.

Bei Bäumen können viele dieser Merkmale erst nach vielen Jahren, teilweise Jahrzehnten untersucht werden, sodass die Forstpflanzenzüchtung nach „neuen Werkzeugen“ „ruft“, mit deren Hilfe diese Merkmale bereits zu einem frühen Zeitpunkt während der Individualentwicklung, im Idealfall bereits im Sämlingsstadium, bestimmt werden können. Diese so genannte Marker-gestützte Züchtung („SMART-Breeding“) [17] erfordert allerdings zunächst die Identifikation der für die Merkmalsausbildung und -ausprägung bedeutenden Gene. In diesen Genen werden dann genetische Unterschiede (z.B. „single nucleotide polymorphisms“, die „SNPs“) identifiziert, die in Nachkommenchaften oder natürlichen Populationen eine Assoziation zu den „gewünschten“ züchterisch relevanten Merkmalen zeigen.

Die Selektion von Individuen/Nachkommen, die den „gewünschten“ molekularen Marker (z. B. SNPs) haben, führt dann im Ergebnis zu besser angepasstem und anpassungsfähigem Vermehrungsgut.

Ausblick

Witterungsextreme und Klimawandel stellen höchste Anforderungen an das individuelle Selbstregulationsvermögen und die Fitness von Baumindividuen und Populationen. Gerade bei den Nadelbaumarten besteht bei gestiegenem Holzverbrauch und abnehmender Nadelholzwaldfläche in Deutschland die Notwendigkeit der waldbaulichen Risikominderung [18, 19]. Da mithilfe der Genomforschung Stress-toleranzen gegen Witterungseinflüsse gezielt entdeckt werden können, würde die Entwicklung und Erprobung oben genannter molekularer Marker auch Handlungsempfehlungen ermöglichen, die das waldbauliche Risiko mindern helfen.

Insgesamt zeigt sich in den Genomen der Waldbäume ein Informationsschatz, dessen Bedeutung für ökonomische und ökologische Anwendungen heute nur erahnt werden kann. Wie nützlich wäre ein fundamentales Verständnis der Gene, die für die Holzbildung verantwortlich sind, die der Abwehr von Schädlingen und Krankheiten dienen, die schnelles Wachstum bedingen, die eine effiziente Kohlenstoffsequestrierung ermöglichen oder die eine Entwicklung biobasierter erneuerbarer Produkte und die Gewinnung schadstoffarmer Energie aus Biomasse erlauben? Deutschland allerdings spielt bei diesen Themen ohne eine erheblich gesteigerte Forschungsförderung leider keine oder eine nur untergeordnete Rolle!

Literaturhinweise:

- [1] <http://phys.org/news/2013-05-team-norway-spruce-genome-sequence.html>. [2] The Arabidopsis Genome Initiative (2000): Nature 14:408(6814): 796-815. [3] NYSTEDT, B.; STREET, N. R.; WETTERBOM A, ZUCCOLO, A. et al. (2013): Nature, published online 22 May 2013, doi:10.1038/nature12211. [4] TUSKAN GA, DIFAZIO S, JANSSEN, S.; BOHLMANN, J. et al. (2006) Science 313: 1596-604. [5] FLADUNG, M. (2005): AFZ-DerWald 5/2005: 248-252. [6] MURRAY, B. G.; LEITCH, I. J.; BENNETT, M. D. (2004): http://www.kew.org/cvalues/Release_3.0, accessed Dec 2004. [7] MACKAY, J.; DEAN, J. F. D.; PLOMION, C.; PETERSON, D. G. et al. (2012) Plant Mol Biol 80: 555-569. [8] <http://www.phytozome.net/poplar.php#C>. [9] <http://www.phytozome.net/eucalyptus.php#B>. [10] <http://www.phytozome.net/arabidopsis.php#B>. [11] LU, S.; SUN, Y.-H.; SHI, R.; CLARK, C.; et al. (2005): The Plant Cell 17: 2186-2203. [12] DOLGOSHEINA, E.V.; MORIN, R. D.; AKSAY, G.; SAHINALP, S.C. et al. (2008): RNA 14: 1508-1515. [13] MACKAY, J. J.; DEAN, J. F. D. (2011): In: Plomion, C.; Bousquet, J.; Kole, C. (Eds) Ednbridge Science Publishers and CRC Press, New York, Seiten 323-357. [14] COWLEY, M.; OAKLEY, R. J. (2013) PLoS Genet. 9(1):e1003234. [15] BIROL, I.; RAYMOND, A.; JACKMAN, S. D.; Pleasance, S et al. (2013) Bioinformatics Advance Access published May 22, 2013, doi:10.1093/bioinformatics/btt178. [16] WHEELER, N.; NEALE, D. B. Available at: <http://www.extension.org/pages/67931/reference-genome-sequencing-conifer-genomics-module-17> (verified April 29, 2013). [17] FLADUNG, M.; GEBHARDT, K. (2010). Forst- und Holz 65: 37-40. [18] SPELLMANN, H.; ALBERT, M.; SCHMIDT, M.; SUTMÖLLER, J.; OVERBECK, M. (2011) AFZ-DerWald 11/2011: 19-23. [19] SPELLMANN, H. (2013) AFZ-DerWald 9/2013:10-15.

Glossar:

- **cDNA:** komplementäre DNA, die mittels des Enzyms Reverse-Transkriptase aus RNA synthetisiert wird
- **DNA:** Desoxyribonukleinsäure
- **Epigenetik:** Vererbung von Eigenschaften, die nicht im DNA-Code festgelegt sind
- **ESTs:** expressed sequence tags (Teil einer cDNA-Sequenz)
- **Exon:** exprimierte Region eines eukaryotischen Gens
- **Gen:** Erbanlage (chemisch, Abschnitt der DNA). Meist wird mit Gen ein Abschnitt der DNA bezeichnet, der die Grundinformationen zur Herstellung einer biologisch aktiven RNA enthält. Es gibt verschiedene Arten von RNAs, die bekannteste ist die mRNA, aus der im Prozess der Translation Proteine gebildet werden. Die Proteine übernehmen im Körper jeweils ganz spezifische Funktionen, die an der Ausprägung von Merkmalen beteiligt sind. Allgemein werden Gene als Erbanlage bezeichnet, da sie die Träger von Erbinformation sind und durch Reproduktion an die Nachkommen weitergegeben werden.
- **Genexpression:** Bildung eines von einem Gen kodierten Genproduktes, vor allem von RNA-Molekülen und Proteinen. Die Genexpression besteht aus Transkription (Bildung von RNA-Molekülen) und gegebenenfalls Translation (Bildung von Proteinmolekülen aus mRNA, messenger RNA) sowie deren Regulationsmechanismen.
- **Genom:** Gesamtheit der vererbaren Information einer Art
- **Intron:** nichtexprimierte Region eines eukaryotischen Gens
- **microRNA (miRNA oder miR):** kurze (21 bis 23 Nukleotide lang), hoch konservierte, nicht kodierende RNAs, die eine wichtige Rolle in dem komplexen Netzwerk der Genregulation, insbesondere beim Gen-Silencing spielen. MicroRNAs regulieren die Genexpression hochspezifisch auf der post-transkriptionalen Ebene. miRNAs werden durch Gene kodiert.
- **Mutation:** spontane Veränderung des Erbgutes
- **Kohlenstoffsequestrierung:** Kohlenstoff-Bindung
- **RNA:** Ribonukleinsäure
- **small RNAs(sRNAs):** kurze, 21 bzw. 24 Nukleotide kleine RNA-Moleküle
- **SMART:** Abk. für engl. Selection with Markers and Advanced Reproductive Technologies: Selektion mit Markern und Nutzung von anspruchsvollen reproduktiven Techniken
- **SMART-Breeding:** Züchtung unter Einsatz der SMART-Technologie
- **SNPs:** Abk. für engl. Single Nucleotide Polymorphisms: Einzelnukleotidvarianten