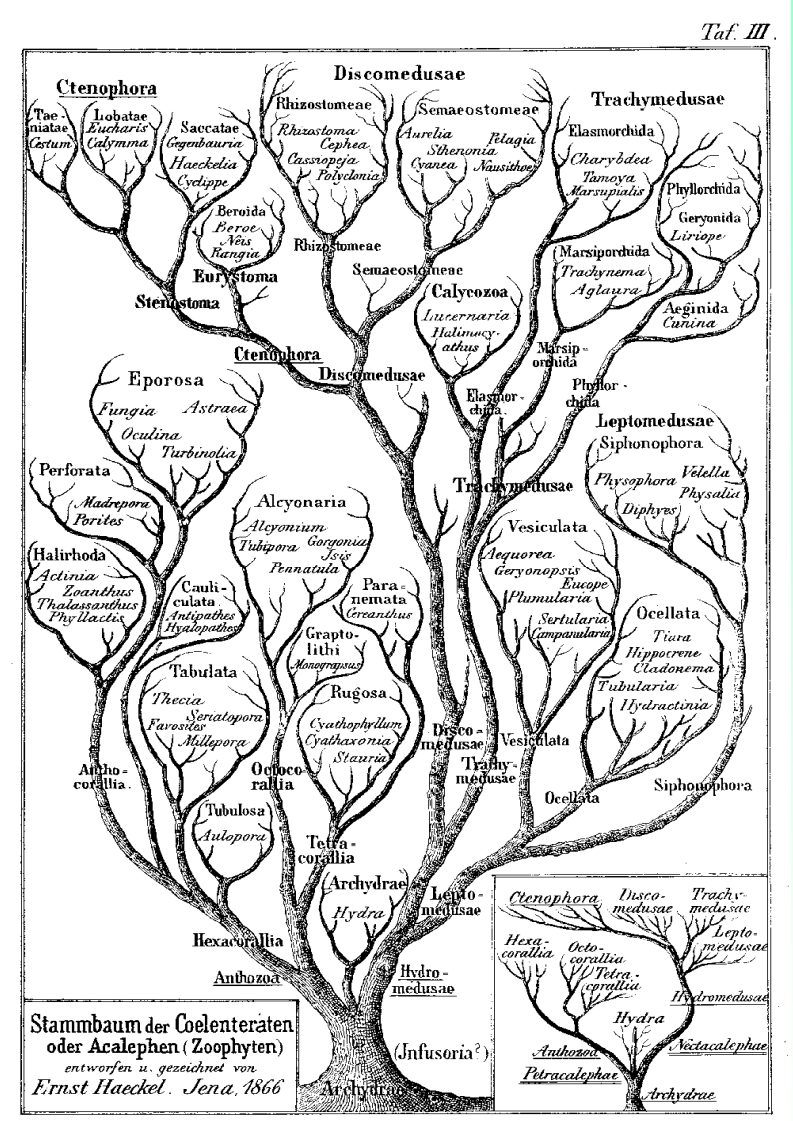
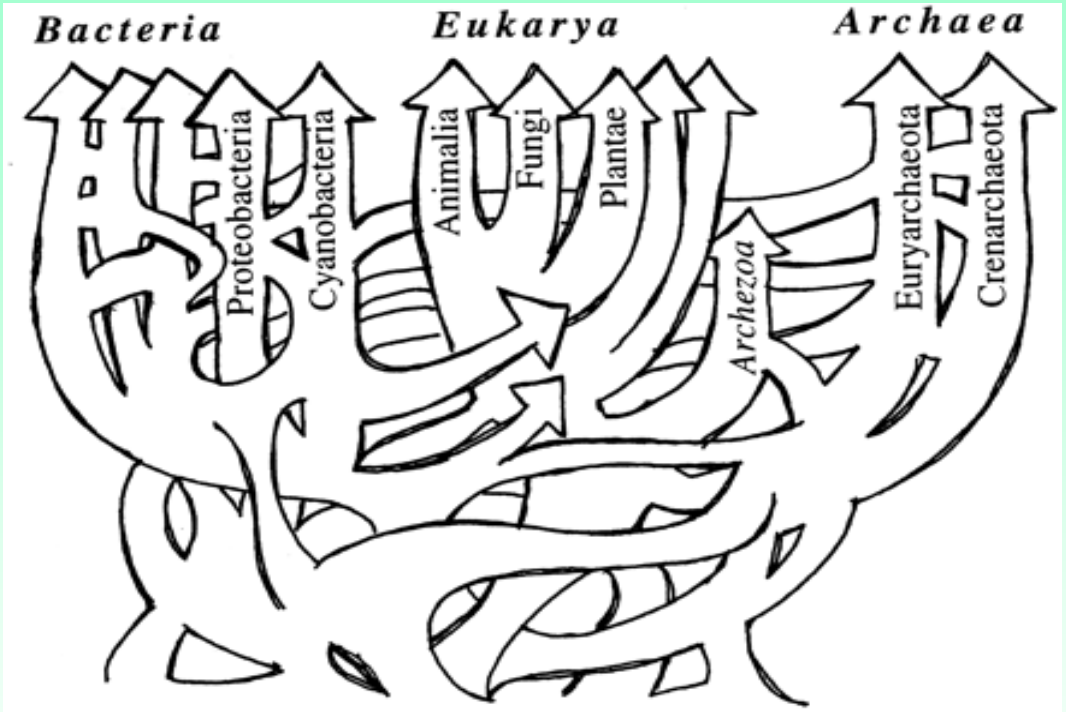
Horizontal Gene Transfer (HGT) : Phylogenetic Prediction and Experimental Test

Yang Zhong

(Fudan University / Tibet University)

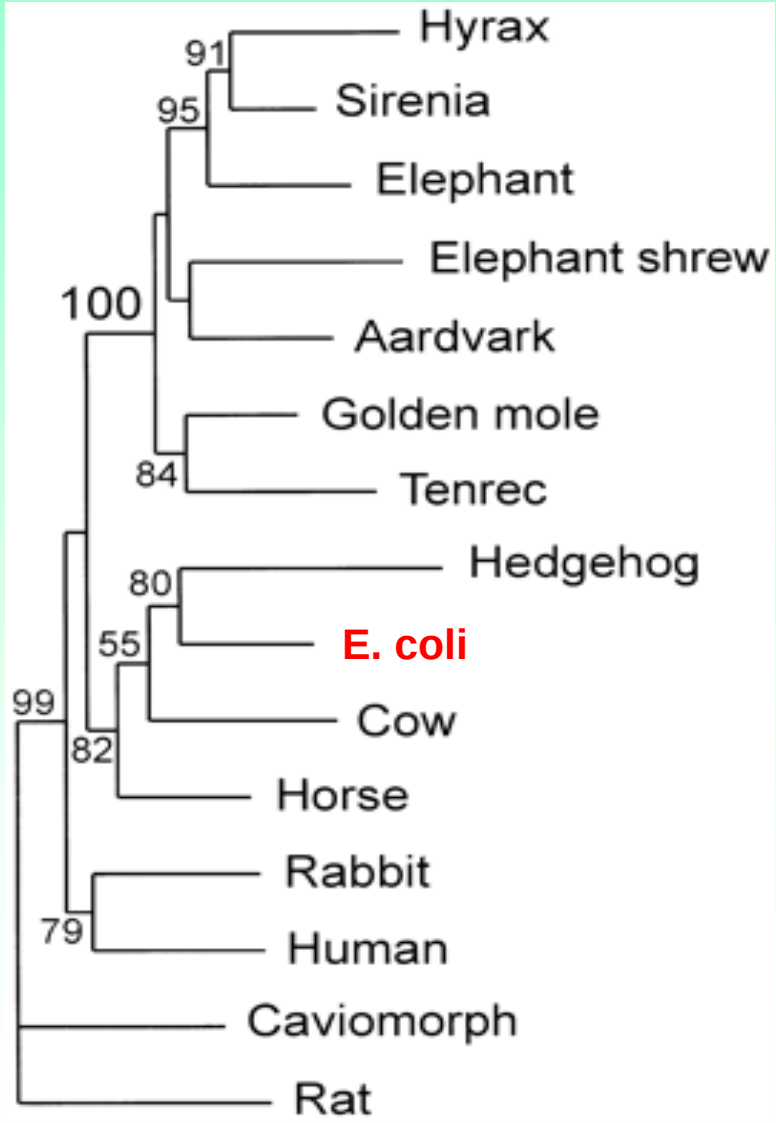


“Tree of Life”

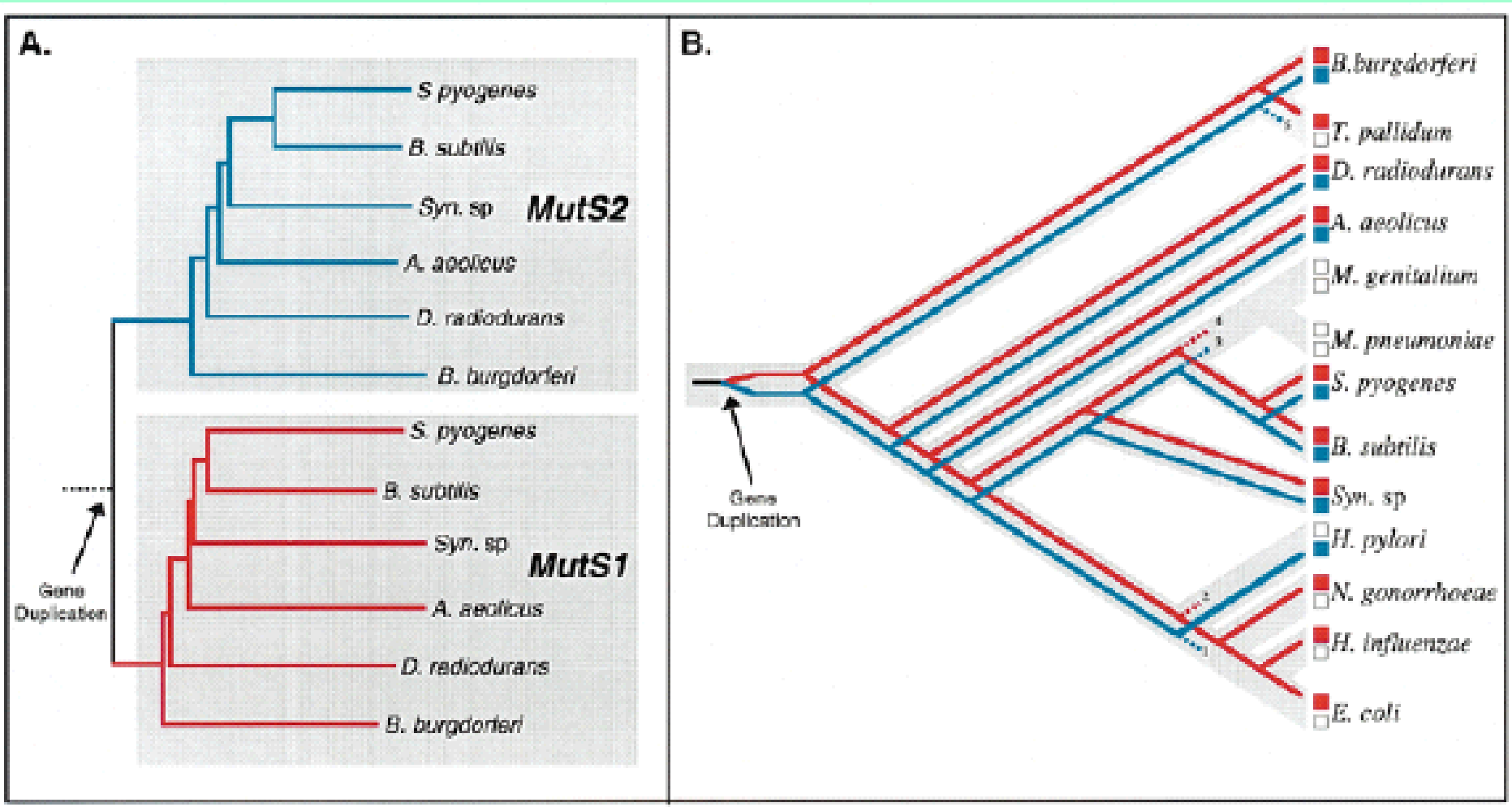


“Network of Life”

HGT vs. Gene loss



Detecting gene loss using the gene trees and species tree



articles

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

* *A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.*

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

Genome Sequencing Centres (Listed in order of total genomic sequence contributed, with a partial list of personnel. A full list of contributors at each centre is available as Supplementary Information.)

Whitehead Institute for Biomedical Research, Center for Genome Research: Eric S. Lander[1]*, Lauren M. Linton[1], Bruce Birren[1]*, Chad Nusbaum[1]*, Michael C. Zody[1]*, Jennifer Baldwin[1], Keri Devon[1], Ken Dewar[1], Michael Doyle[1], William FitzHugh[1]*, Roel Funke[1], Diane Gage[1], Katrina Harris[1], Andrew Heaford[1], John Howland[1], Lisa Kann[1], Jessica Lehoczky[1], Rosie LeVine[1], Paul McEwan[1], Kevin McKernan[1], James Meldrim[1], Jill P. Mesirov[1]*, Cher Miranda[1], William Morris[1], Jerome Naylor[1], Christina Raymond[1], Mark Rosetti[1], Ralph Santos[1], Andrew Sheridan[1], Carrie Sougnez[1], Nicole Stange-Thomann[1], Nikola Stojanovic[1], Aravind Subramanian[1] & Dudley Wyman[1]

The Sanger Centre: Jane Rogers[2], John Sulston[2]*, Rachael Ainscough[2], Stephan Beck[2], David Bentley[2], John Burton[2], Christopher Clee[2], Nigel Carter[2], Alan Coulson[2], Rebecca Deadman[2], Panos Deloukas[2], Andrew Dunham[2], Ian Dunham[2], Richard Durbin[2]*, Lisa French[2], Darren Grafham[2], Simon Gregory[2], Tim Hubbard[2]*, Sean Humphray[2], Adrienne Hunt[2], Matthew Jones[2], Christine Lloyd[2], Amanda McMurray[2], Lucy Matthews[2], Simon Mercer[2], Sarah Milne[2], James C. Mullikin[2]*, Andrew Mungall[2], Robert Plumb[2], Mark Ross[2], Ratna Shownkeen[2] & Sarah Sims[2]

Washington University Genome Sequencing Center: Robert H. Waterston[3]*, Richard K. Wilson[3], LaDeana W. Hillier[3]*, John D. McPherson[3], Marco A. Marra[3], Elaine R. Mardis[3], Lucinda A. Fulton[3], Asif T. Chinwalla[3]*, Kymberlie H. Pepin[3], Warren R. Gish[3], Stephanie L. Chissoe[3], Michael C. Wendl[3], Kim D. Delehaunty[3], Tracie L. Miner[3], Andrew Delehaunty[3], Jason B. Kramer[3], Lisa L. Cook[3], Robert S. Fulton[3], Douglas L. Johnson[3], Patrick J. Minx[3] & Sandra W. Clifton[3]

US DOE Joint Genome Institute: Trevor Hawkins[4], Elbert Branscomb[4], Paul Predki[4], Paul Richardson[4], Sarah Wenning[4], Tom Slezak[4], Norman Doggett[4], Jan-Fang Cheng[4], Anne Olsen[4], Susan Lucas[4], Christopher Elkin[4], Edward Uberbacher[4] & Marvin Frazier[4]

Baylor College of Medicine Human Genome Sequencing Center: Richard A. Gibbs[5]*, Donna M. Muzny[5], Steven E. Scherer[5], John B. Bouck[5]*, Erica J. Sodergren[5], Kim C. Worley[5]*, Catherine M. Rives[5], James H. Gorrell[5], Michael L. Metzker[5], Susan L. Naylor[5], Raju S. Kucherlapati[5], David L. Nelson, & George M. Weinstock[5]

RIKEN Genomic Sciences Center: Yoshiyuki Sakaki[9], Asao Fujiyama[9], Masahira Hattori[9], Tetsushi Yada[9], Atsushi Toyoda[9], Takehiko Itoh[9], Chiharu Kawagoe[9], Hidemi Watanabe[9], Yasushi Totoki[9] & Todd Taylor[9]

Genoscope and CNRS UMR-8030: Jean Weissenbach[10], Roland Heilig[10], William Saurin[10], Francois Artiguenave[10], Philippe Brottier[10], Thomas Bruls[10], Eric Pelletier[10], Catherine Robert[10] & Patrick Wincker[10]

GTC Sequencing Center: Douglas R. Smith[11], Lynn Doucette-Stamm[11], Marc Rubenfield[11], Keith Weinstock[11], Hong Mei Lee[11] & JoAnn Dubois[11]

Department of Genome Analysis, Institute of Molecular Biotechnology: André Rosenthal[12], Matthias Platzer[12], Gerald Nyakatura[12], Stefan Taudien[12] & Andreas Rump[12]

Beijing Genomics Institute/Human Genome Center: Huanming Yang[13], Jun Yu[13], Jian Wang[13], Guyang Huang[14] & Jun Gu[15]

Multimegabase Sequencing Center, The Institute for Systems Biology: Leroy Hood[16], Lee Rowen[16], Anup Madan[16] & Shizen Qin[16]

Stanford Genome Technology Center: Ronald W. Davis[17], Nancy A. Federspiel[17], A. Pia Abola[17] & Michael J. Proctor[17]

Stanford Human Genome Center: Richard M. Myers[18], Jeremy Schmutz[18], Mark Dickson[18], Jane Grimwood[18] & David R. Cox[18]

University of Washington Genome Center: Maynard V. Olson[19], Rajinder Kaul[19] & Christopher Raymond[19]

Department of Molecular Biology, Keio University School of Medicine: Nobuyoshi Shimizu[20], Kazuhiko Kawasaki[20] & Shinsei Minoshima[20]

University of Texas Southwestern Medical Center at Dallas: Glen A. Evans[21]†, Maria Athanasiou[21] & Roger Schultz[21]

University of Oklahoma's Advanced Center for Genome Technology: Bruce A. Roe[22], Feng Chen[22] & Huaqin Pan[22]

Max Planck Institute for Molecular Genetics: Juliane Ramser[23], Hans Lehrach[23] & Richard Reinhardt[23]

Cold Spring Harbor Laboratory, Lita Annenberg Hazen Genome Center: W. Richard McCombie[24], Melissa de la Bastide[24] & Neilay Dedhia[24]

GBF—German Research Centre for Biotechnology: Helmut Blöcker[25], Klaus Hornischer[25] & Gabriele Nordsiek[25]

* **Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings):** Richa Agarwala[26], L. Aravind[26], Jeffrey A. Bailey[27], Alex Bateman[2], Serafim Batzoglou[1], Ewan Birney[28], Peer Bork[29,30], Daniel G. Brown[1], Christopher B. Burge[31], Lorenzo Cerutti[28], Hsiu-Chuan Chen[26], Deanna Church[26], Michele Clamp[2], Richard R. Copley[30], Tobias Doerks[29,30], Sean R. Eddy[32], Evan E. Eichler[27], Terrence S. Furey[33], James Galagan[1], James G. R. Gilbert[2], Cyrus Harmon[34], Yoshihide Hayashizaki[35], David Haussler[36], Henning Hermjakob[28], Karsten Hokamp[37], Wonhee Jang[26], L. Steven Johnson[32], Thomas A. Jones[32], Simon Kasif[38], Arek Kasprzyk[28], Scot Kennedy[39], W. James Kent[40], Paul Kitts[26], Eugene V. Koonin[26], Ian Korf[3], David Kulp[34], Doron Lancet[41], Todd M. Lowe[42], Aoife McLysaght[27], Tarjei Mikkelsen[36], John V. Moran[43], Nicola Mulder[28], Victor J. Pollara[1], Chris P. Ponting[44], Greg Schuler[26], Jörg Schultz[30], Guy Slater[28], Arian F. A. Smit[45], Elia Stupka[28], Joseph Szustakowki[38], Danielle Thierry-Mieg[26], Jean Thierry-Mieg[26], Lukas Wagner[26], John Wallis[3], Raymond Wheeler[34], Alan Williams[34], Yuri I. Wolf[26], Kenneth H. Wolfe[27], Shiaw-Pyng Yang[3] & Ru-Fang Yeh[31]

Scientific management: National Human Genome Research Institute, US National Institutes of Health: Francis Collins[46]*, Mark S. Guyer[46], Jane Peterson[46], Adam Felsenfeld[46]* & Kris A. Wetterstrand[46]; **Office of Science, US Department of Energy:** Aristides Patrinos[47]; **The Wellcome Trust:** Michael J. Morgan[48]

Hundreds of genes appear to have resulted from horizontal gene transfer from bacteria…

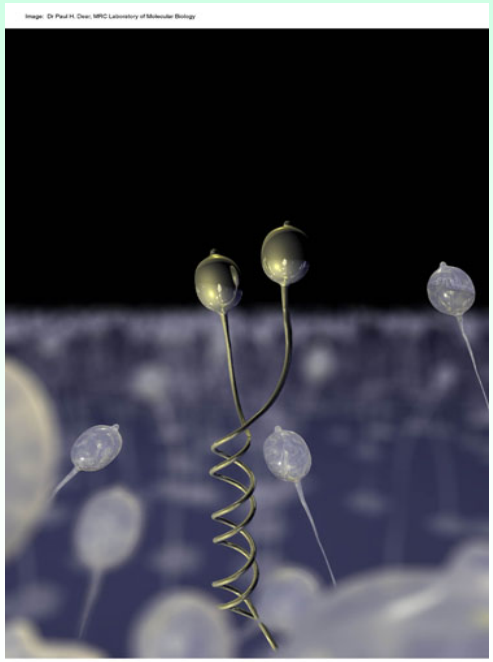
Nature 2001 409:860-921

Genomics

Genes lost during evolution

One of the main conclusions presented by the International Human Genome Sequencing Consortium is that "hundreds of genes appear to have resulted from horizontal gene transfer from bacteria at some point in the vertebrate lineage"[1]. We noticed that a significant proportion of these human genes have closely related orthologues in the primitive eukaryote *Dictyostelium*. This observation supports independent gene loss in multiple lineages (worm, fly, yeast, plants) rather than horizontal gene transfer from bacteria.

The human genome sequence revealed 113 genes that share a high degree of identity with bacterial genes, but are absent in the completely sequenced genomes of *Caenorhabditis elegans*, *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Arabidopsis thaliana*[1]. Do these genes represent examples of horizontal gene transfer from bacteria to the vertebrate lineage, or were they present in both prokaryotes and early eukaryotes, but subsequently lost from all non-vertebrate eukaryotic lineages? Although this latter possibility may seem unlikely, we recently identified a gene in *Dictyostelium* that is clearly an orthologue of the gene that encodes soluble



Within the group of 113 genes proposed to have entered the human genome by horizontal gene transfer from bacteria, we have identified at least 11 that probably arose through normal evolution with gene loss in several lineages, suggesting that gene loss is not a rare event. With several ongoing genomic sequencing projects for lower eukaryotes, it will be interesting to see how many genes have truly undergone horizontal transfer.

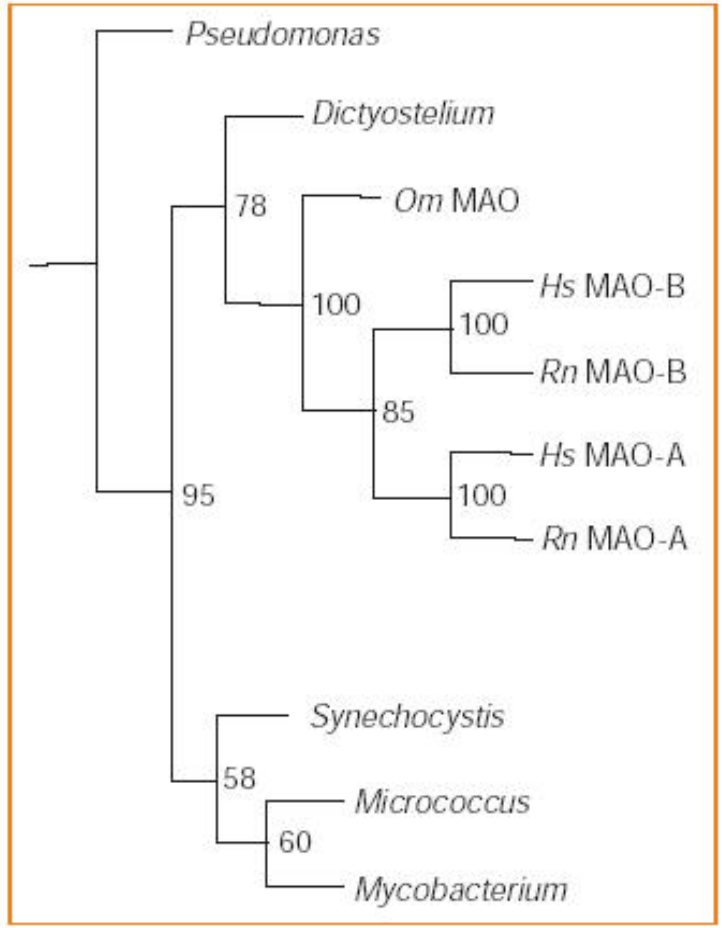


Figure 1 Phylogenetic analysis of monoamine oxidase (MAO). Numbers indicate values of bootstrap analysis ($n = 100$). Hs, *Homo sapiens*; Rn, *Rattus norvegicus* (rat); Om, *Oncorhynchus mykiss* (rainbow trout).

Whole genome of
Dictyostelium discoideum

Nature, 2001 411:1013-1014

HGT is common to organisms at “higher” levels

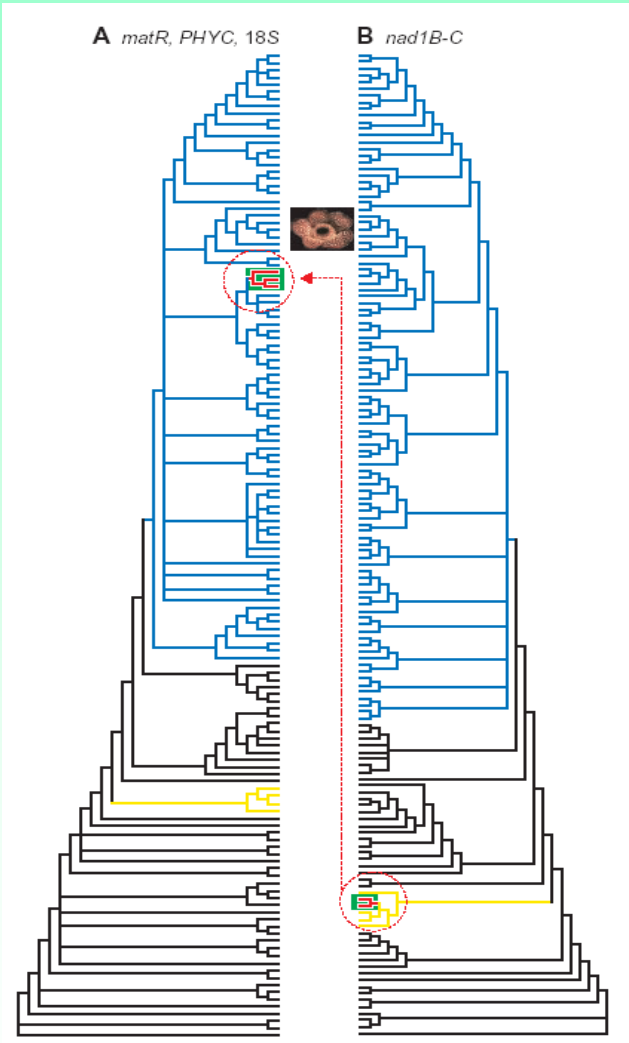
Rafflesia is a parasitic flowering plant, and has no stems, leaves or true roots

Host-to-Parasite Gene Transfer in Flowering Plants: Phylogenetic Evidence from Malpighiales

Charles C. Davis[1]* and Kenneth J. Wurdack[2]

Horizontal gene transfer (HGT) between sexually unrelated species has recently been documented for higher plants, but mechanistic explanations for HGTs have remained speculative. We show that a parasitic relationship may facilitate HGT between flowering plants. The endophytic parasites Rafflesiaceae are placed in the diverse order Malpighiales. Our multigene phylogenetic analyses of Malpighiales show that mitochrodrial (*matR*) and nuclear loci (18*S* ribosomal DNA and *PHYC*) place Rafflesiaceae in Malpighiales, perhaps near Ochnaceae/Clusiaceae. Mitochondrial *nad1B-C*, however, groups them within Vitaceae, near their obligate host *Tetrastigma*. These discordant phylogenetic hypotheses strongly suggest that part of the mitochondrial genome in Rafflesiaceae was acquired via HGT from their hosts.

Two conflicting hypotheses about the phylogenetic placement of Rafflesiaceae



Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*

Ulfar Bergthorsson†, Aaron O. Richardson, Gregory J. Young, Leslie R. Goertzen‡, and Jeffrey D. Palmer§

Department of Biology, Indiana University, Bloomington, IN 47405-3700

Contributed by Jeffrey D. Palmer, November 9, 2004

Several plants are known to have acquired a single mitochondrial gene by horizontal gene transfer (HGT), but whether these or any other plants have acquired many foreign genes is entirely unclear. To address this question, we focused on *Amborella trichopoda*, because it was already known to possess one horizontally acquired gene and because it was found in preliminary analyses to contain several more. We comprehensively sequenced the mitochondrial protein gene set of *Amborella*, sequenced a variable number of mitochondrial genes from 28 other diverse land plants, and conducted phylogenetic analyses of these sequences plus those already available, including the five sequenced mitochondrial genomes of angiosperms. Results indicate that *Amborella* has acquired one or more copies of 20 of its 31 known mitochondrial protein genes from other land plants, for a total of 26 foreign genes, whereas no evidence for HGT was found in the five sequenced genomes. Most of the *Amborella* transfers are from other angiosperms (especially eudicots), whereas others are from nonangiosperms, including six striking cases of transfer from (at least three different) moss donors. Most of the transferred genes are intact, consistent with functionality and/or recency of transfer. *Amborella* mtDNA has sustained proportionately more HGT than any other eukaryotic, or perhaps even prokaryotic, genome yet examined.

Materials and Methods

We used primers for conserved regions of angiosperm mitochondrial genes in an attempt to PCR-amplify and sequence all mitochondrial protein genes from *A. trichopoda* (primer sequences available on request). Many *Amborella* reactions produced multiple bands, heterogeneous sequence, or unreadable sequence; these were cloned, and multiple (usually eight) clones were sequenced. This process yielded portions of 27 genes. We then used PCR to amplify and sequence as many of these 27 genes as possible, plus the four genes already sequenced from *Amborella* mtDNA, from 13 other angiosperms (see Fig. 5, which is published as supporting information on the PNAS web site, for taxa and sources) and three gymnosperms. For each of these plants, we carried out 80 PCRs with conserved mitochondrial primers. Selected genes were amplified and sequenced from 12 additional nonangiosperms. PCR was performed under the following conditions: 95°C for 2 min, 35 cycles of 95°C for 30 s, 55° or 52°C for 30 s, 72°C for 2 min, and 72°C for 5 min. PCR products were cleaned by using 2 μl of ExoSAP-IT (United States Biochemical). Sequences were generated by using an ABI 3730 (Applied Biosystems). Sequence traces were assembled and trimmed by using CODONCODE ALIGNER 1.3.2.

Sequences were aligned by using either BIOEDIT or SE-AL

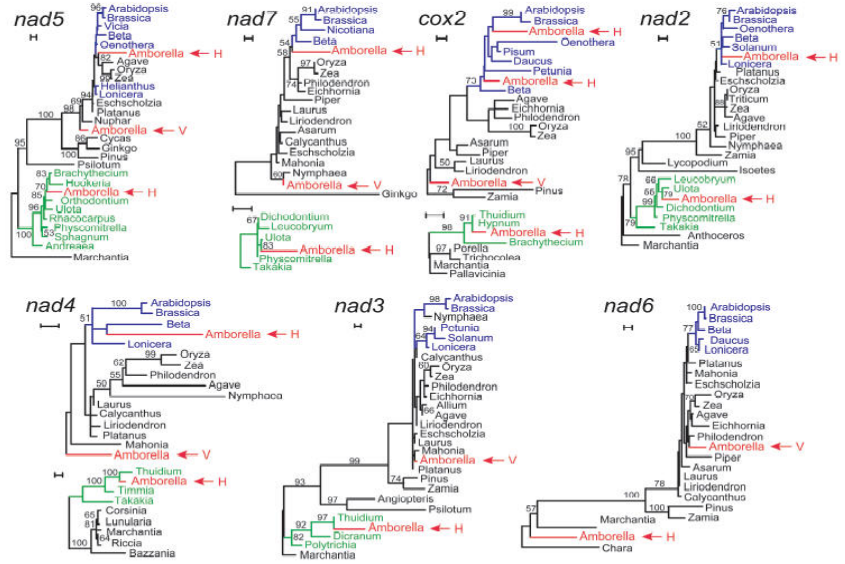


Fig. 1. Phylogenetic evidence for horizontal acquisition of genes from mosses and angiosperms in *Amborella*. Shown are ML trees. Bootstrap values (100 ML replicates) >50% are shown. H and V indicate *Amborella* genes of putatively horizontal and vertical transmission, respectively. *Amborella* genes are in red, core eudicot genes are in blue (basal eudicots commonly included are *Platanus*, *Eschscholzia*, and *Mahonia*), and moss genes are in green. Note that for *nad7*, *cox2*, and *nad4*, seed and nonseed plants were analyzed separately. Scale bars correspond to 0.01 substitutions per site.

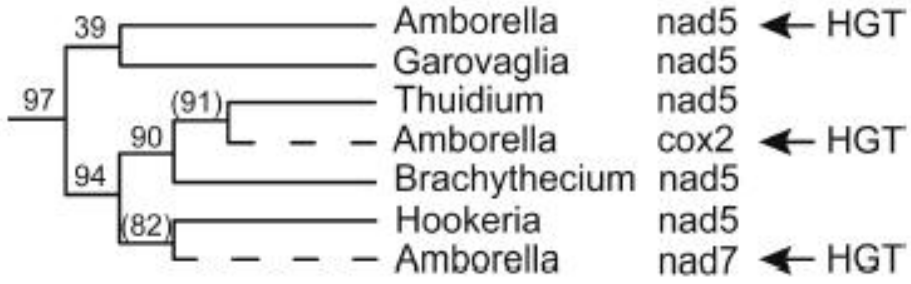


Fig. 2. *Amborella* acquired three genes from different moss donors. The solid parts of the cladogram and nonparenthetical bootstrap values are from the *nad5* intron phylogeny of Fig. 6. The dashed lines and other bootstrap values indicate the relationship to the indicated mosses of the moss-derived *cox2* and *nad7* genes of *Amborella*, as per the *cox2* gene tree of Fig. 1 and the *nad7* intron tree of Fig. 6.

Fig. 4. *A. trichopoda* leaf from a cloud forest at Massif de l'Aoupinié (Province Nord in New Caledonia) at 801 m altitude. Note the greenish bryophyte (liverwort) growth covering the leaf tip, and the small spots of lichens and other epiphytes elsewhere on the leaf. Photograph courtesy of Sean Graham, Centre for Plant Research, University of British Columbia, Vancouver.

Plant genetics

Gene transfer from parasitic to host plants

Plant mitochondrial genes are transmitted horizontally across mating barriers with surprising frequency, but the mechanism of transfer is unclear[1,2]. Here we describe two new cases of horizontal gene transfer, from parasitic flowering plants to their host flowering plants, and present phylogenetic and biogeographic evidence that this occurred as a result of direct physical contact between the two. Our findings complement the discovery that genes can be transferred in the opposite direction, from host to parasite plant[3].

Figure 2 A parasitic dodder (*Cuscuta californica*) in flower, with its haustoria penetrating a host tomato plant.

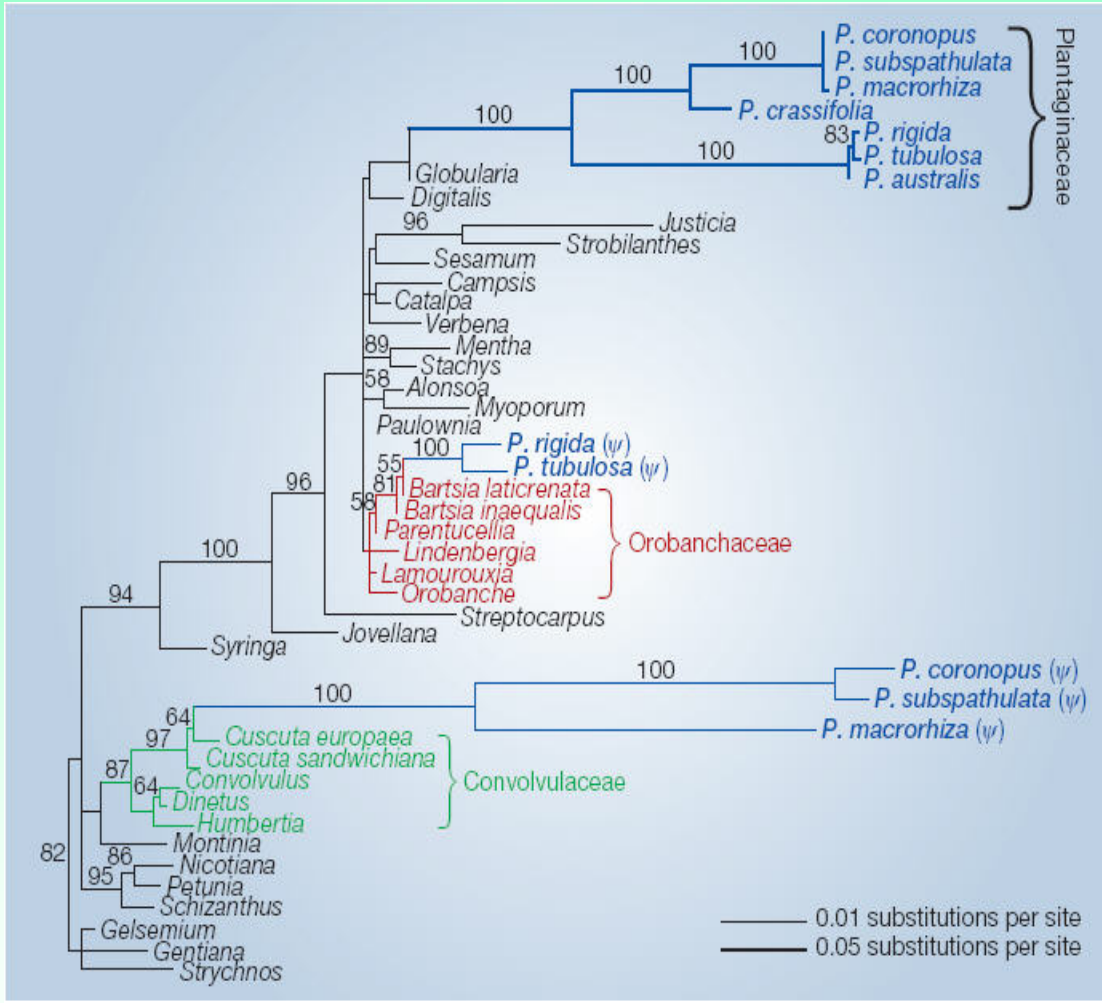


Figure 1 Phylogenetic evidence for two horizontal transfer events of the gene *atp1* into *Plantago* (blue). Seven *Plantago atp1* genes at the top of the maximum-likelihood tree are intact, vertically transmitted, and rapidly evolving (scale reduced by 80%). The other two sets of *Plantago atp1* genes are pseudogenes (ψ) acquired from parasitic plants in the Orobanchaceae (red) and Convolvulaceae (green). Bootstrap values of over 50% are shown. For methods, see supplementary information.

Horizontal gene transfer of the algal nuclear gene *psbO* to the photosynthetic sea slug *Elysia chlorotica*

Mary E. Rumpho[a,1], Jared M. Worful[a], Jungho Lee[b], Krishna Kannan[a], Mary S. Tyler[c], Debashish Bhattacharya[d], Ahmed Moustafa[d], and James R. Manhart[e]

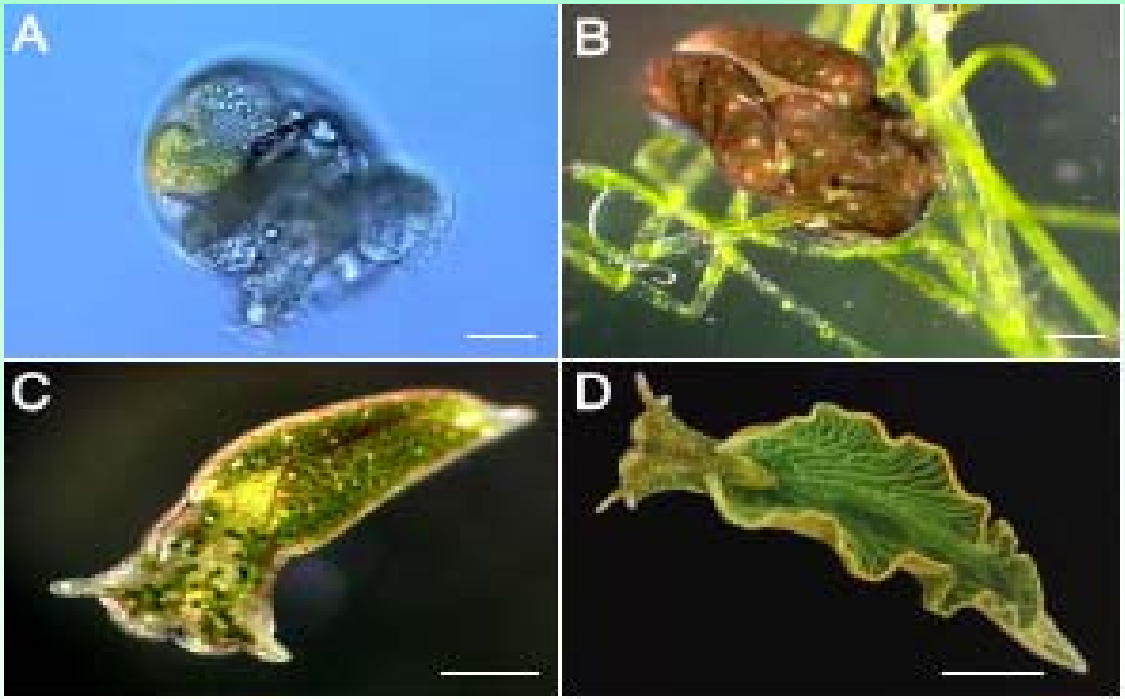
[a]Department of Biochemistry, Microbiology, and Molecular Biology, University of Maine, Orono, ME 04469; [b]Green Plant Institute, Seoul National University, Gwonseon, Suwon, Gyeonggi 441-853, Korea; [c]School of Biology and Ecology, University of Maine, Orono, ME 04469; [d]Department of Biological Sciences and the Roy J. Carver Center for Comparative Genomics, Interdisciplinary Program in Genetics, University of Iowa, Iowa City, IA 52242-1324; and [e]Department of Biology, Texas A&M University, College Station, TX 77843

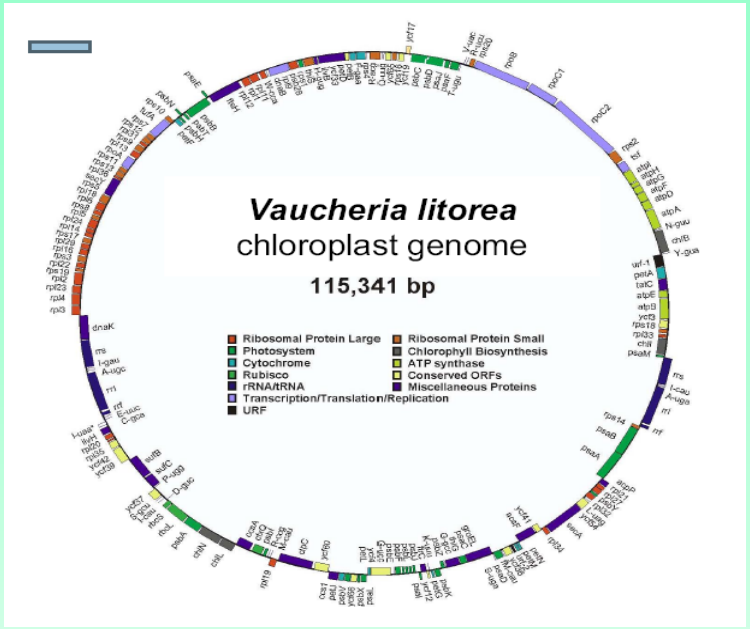
Edited by Lynn Margulis, University of Massachusetts, Amherst, MA, and approved September 17, 2008 (received for review June 9, 2008)

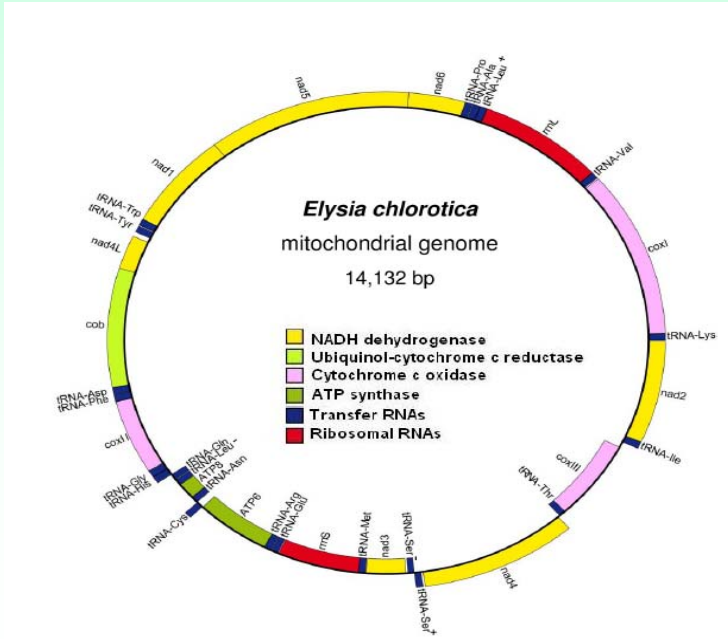
The sea slug *Elysia chlorotica* acquires plastids by ingestion of its algal food source *Vaucheria litorea*. Organelles are sequestered in the mollusc's digestive epithelium, where they photosynthesize for months in the absence of algal nucleocytoplasm. This is perplexing because plastid metabolism depends on the nuclear genome for >90% of the needed proteins. Two possible explanations for the persistence of photosynthesis in the sea slug are (*i*) the ability of *V. litorea* plastids to retain genetic autonomy and/or (*ii*) more likely, the mollusc provides the essential plastid proteins. Under the latter scenario, genes supporting photosynthesis have been acquired by the animal via horizontal gene transfer and the encoded proteins are retargeted to the plastid. We sequenced the plastid genome and confirmed that it lacks the full complement of genes required for photosynthesis. In support of the second scenario, we demonstrated that a nuclear gene of oxygenic photosynthesis, *psbO*, is expressed in the sea slug and has integrated into the germline. The source of *psbO* in the sea slug is *V. litorea* because this sequence is identical from the predator and prey genomes. Evidence that the transferred gene has integrated into sea slug nuclear DNA comes from the finding of a highly diverged *psbO* 3′ flanking sequence in the algal and mollusc nuclear homologues and gene absence from the mitochondrial genome of *E. chlorotica*. We demonstrate that foreign organelle retention generates metabolic novelty ("green animals") and is explained by anastomosis of distinct branches of the tree of life driven by predation and horizontal gene transfer.

20). Most of these latter examples are associated with parasitism or phagotrophy, including the elegant studies of HGT from the α-proteobacteria *Wolbachia* to insects and nematodes (16–18), and the finding of rhizobial-like genes in plant parasitic nematodes (19, 20). The exchange of genetic material between two eukaryotes is extremely rare, or at least not well documented to date. The best-studied cases include the transfer of mitochondrial DNA from achlorophyllous or epiphytic plants to the mitochondrial genome (mtDNA) of their closely related photosynthetic hosts (21), the exchange of transposons between two animal (22) or two plant (23) species, and the presence of plant genes in plant parasitic nematodes (in addition to the rhizobial genes discussed previously), which are hypothesized to be "defense" genes whose products protect the parasite from host detection (20).

The sacoglossan mollusc (sea slug) *Elysia chlorotica* represents a unique model system to study the potential for interdomain HGT between two multicellular eukaryotes—in this case, from a filamentous secondary (heterokont) alga (*Vaucheria litorea*) to a mollusc. This emerald green sea slug owes its coloring and photosynthetic ability to plastids acquired during herbivorous feeding (24–29). The plastids do not undergo division in the sea slug and are sequestered intracellularly in cells lining the finely divided digestive diverticula. The plastids continue to carry out photosynthesis, providing the sea slug with energy and carbon during its approximately 10-month life span (27, 28). Long-term plastid activity continues despite the absence of algal nuclei (27, 29), and hence a source of nuclear-encoded plastid-targeted


A
B
C
D


Vaucheria litorea
chloroplast genome
115,341 bp
Ribosomal Protein Large
Photosystem
Cytochrome
Rubisco
rRNA/tRNA
Transcription/Translation/Replication
URF
Ribosomal Protein Small
Chlorophyll Biosynthesis
ATP synthase
Conserved ORFs
Miscellaneous Proteins


Elysia chlorotica
mitochondrial genome
14,132 bp
NADH dehydrogenase
Ubiquinol-cytochrome c reductase
Cytochrome c oxidase
ATP synthase
Transfer RNAs
Ribosomal RNAs

Cistanche deserticola

Haloxylon ammodendron

Cistanche deserticola ?

OPEN ACCESS Freely available online

PLOS ONE

Complete Chloroplast Genome Sequence of Holoparasite *Cistanche deserticola* (Orobanchaceae) Reveals Gene Loss and Horizontal Gene Transfer from Its Host *Haloxylon ammodendron* (Chenopodiaceae)

Xi Li[1☯], Ti-Cao Zhang[1☯], Qin Qiao[1], Zhumei Ren[2], Jiayuan Zhao[1], Takahiro Yonezawa[1], Masami Hasegawa[1], M. James C Crabbe[3], Jianqiang Li[4]*, Yang Zhong[1,5]*

1 Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, School of Life Sciences, Fudan University, Shanghai, China, **2** College of Life Science and Technology, Shanxi University, Taiyuan, China, **3** Faculty of Creative Arts, Technologies and Science, Institute of Biomedical, Environmental Science and Technology, University of Bedfordshire, Luton, United Kingdom, **4** Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China, **5** Institute of Biodiversity Science and Geobiology, Tibet University, Lhasa, China

Abstract

Background: The central function of chloroplasts is to carry out photosynthesis, and its gene content and structure are highly conserved across land plants. Parasitic plants, which have reduced photosynthetic ability, suffer gene losses from the chloroplast (cp) genome accompanied by the relaxation of selective constraints. Compared with the rapid rise in the number of cp genome sequences of photosynthetic organisms, there are limited data sets from parasitic plants.

Principal Findings/Significance: Here we report the complete sequence of the cp genome of *Cistanche deserticola*, a holoparasitic desert species belonging to the family Orobanchaceae. The cp genome of *C. deserticola* is greatly reduced both in size (102,657 bp) and in gene content, indicating that all genes required for photosynthesis suffer from gene loss and pseudogenization, except for *psb*M. The striking difference from other holoparasitic plants is that it retains almost a full set of tRNA genes, and it has lower *dN/dS* for most genes than another close holoparasitic plant, *E. virginiana*, suggesting that *Cistanche deserticola* has undergone fewer losses, either due to a reduced level of holoparasitism, or to a recent switch to this life history. We also found that the *rpo*C2 gene was present in two copies within *C. deserticola*. Its own copy has much shortened and turned out to be a pseudogene. Another copy, which was not located in its cp genome, was a homolog of the host plant, *Haloxylon ammodendron* (Chenopodiaceae), suggesting that it was acquired from its host via a horizontal gene transfer.

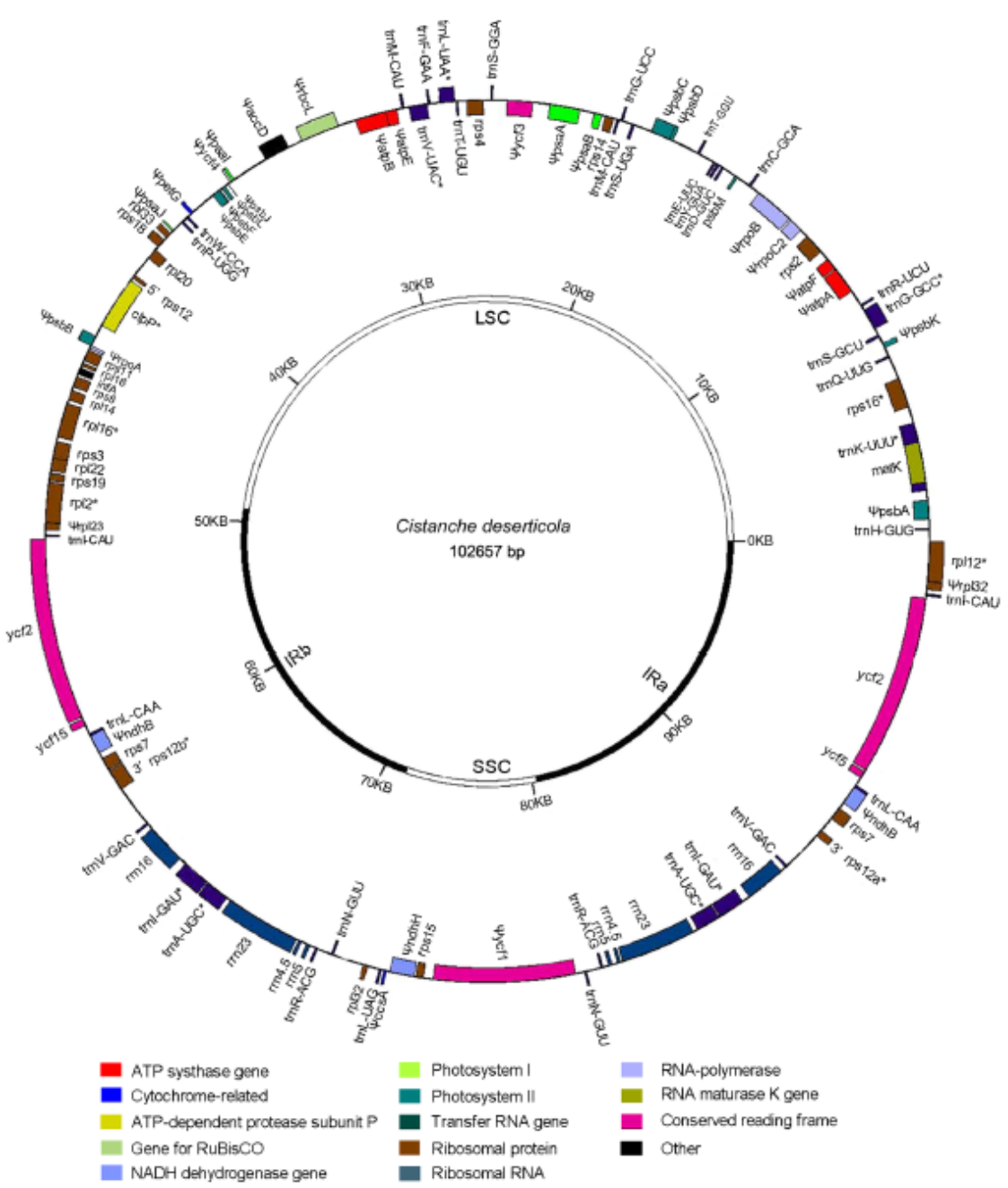


Figure 1. Gene Maps of the plastid chromosomes of *C. deserticola*. Genes shown inside the circle are transcribed clockwise, those outside the circle are transcribed counterclockwise. The large single copy region (LSC) and the small single copy region (SSC) are separated by two inverted repeats (IRa and IRb). Asterisks indicate intron containing genes. Pseudogenes are marked by Ψ.
doi:10.1371/journal.pone.0058747.g001

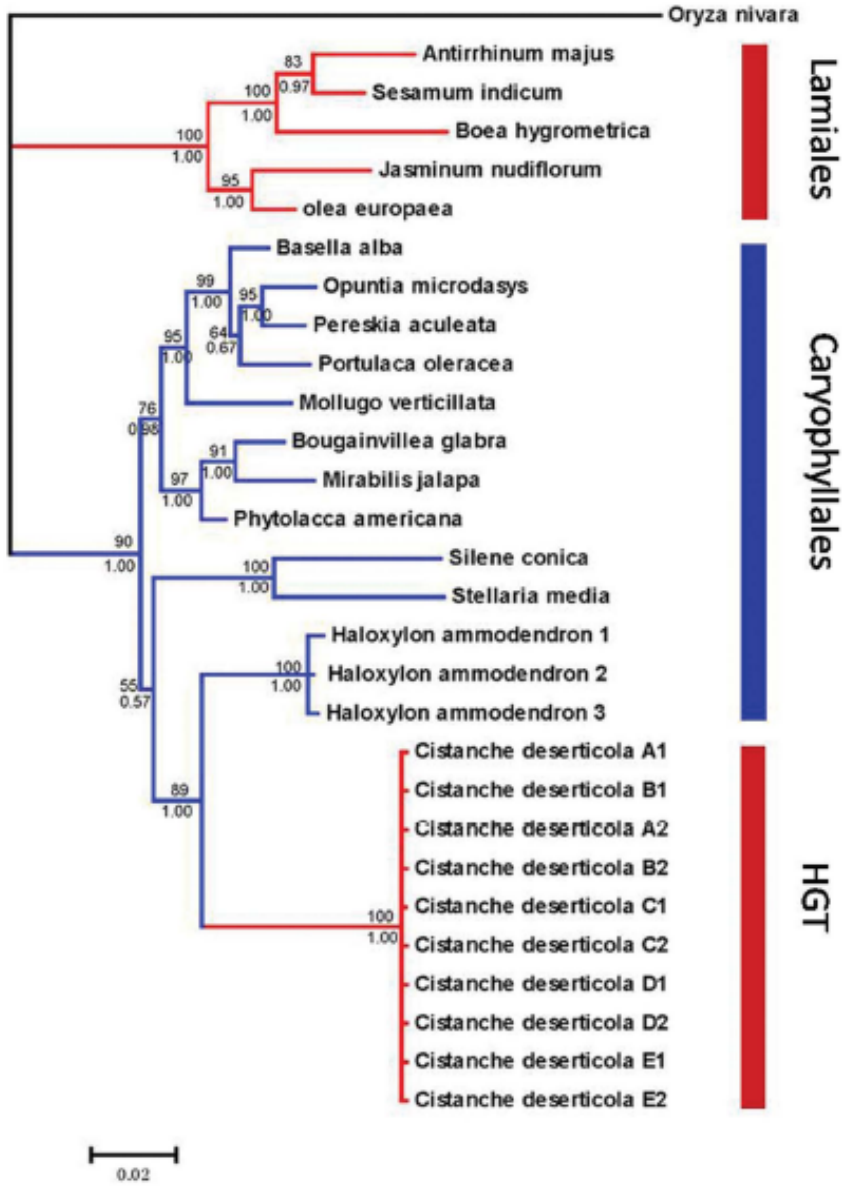


Figure 3. Phylogenetic evidence for horizontal gene transfer of the plastid rpoC2 from *Haloxylon ammodendron* to *Cistanche deserticola*. Lamiales are coloured in red, and Caryophyllales are coloured in blue. While the ten *C. deserticola* sequences involved in horizontal gene transfer are coloured in red. Numbers at nodes are posterior probabilities >0.60 and maximum likelihood bootstrap values >60. The Genebank number: *Oryza nivara*, NC_005973; *Antirrhinum indicum*, GQ997028; *Sesamum indicum*, NC_016433; *Boea hygrometrica*, NC_016468; *Jasminum nudiflorum*, NC_008407; *Olea europaea*, NC_013707; *Basella alba*, HQ843359; *Opuntia microdasys*, HQ843375; *Pereskia aculeata*, HQ843376; *Portulaca oleracea*, HQ843380; Mollugo verticillata, HQ843373; *Bougainvillea glabra*, HQ843360; *Mirabilis jalapa*, HQ843372; *Phytolacca americana*, HQ843378; *Celosia cristata*, HQ843361; *Spinacia oleracea*, NC_002202; *Silene conica*, NC_016729; *Stellaria media*, HQ843386; *Cistanche deserticola* (HGT), KC543998.

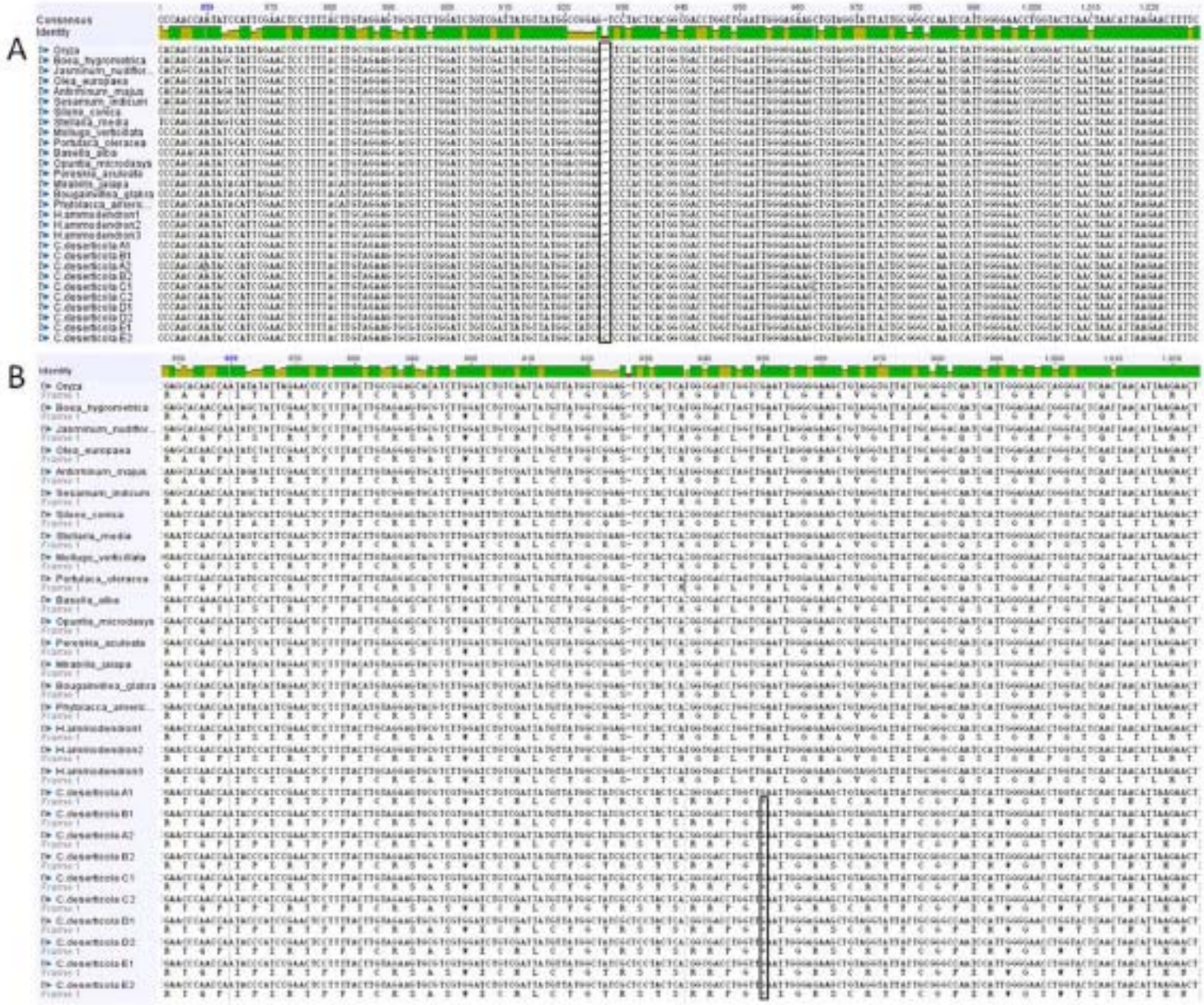


Figure 4. Position of inserted cytosine within the transferred rpoC2 gene. (A) Alignment of the nucleotide sequences of transferred rpoC2 gene amplified from parasite and host with intact open reading frames of other related species. The inserted cytosine was labeled with colored vertical lines. (B) Inserted cytosine resulted in followed premature termination codon in the transferred rpoC2 in *Cistanche deserticola*.
doi:10.1371/journal.pone.0058747.g004

Rhus gall aphids and their host plants: a case study

Rhus gall aphids and their lifecycles

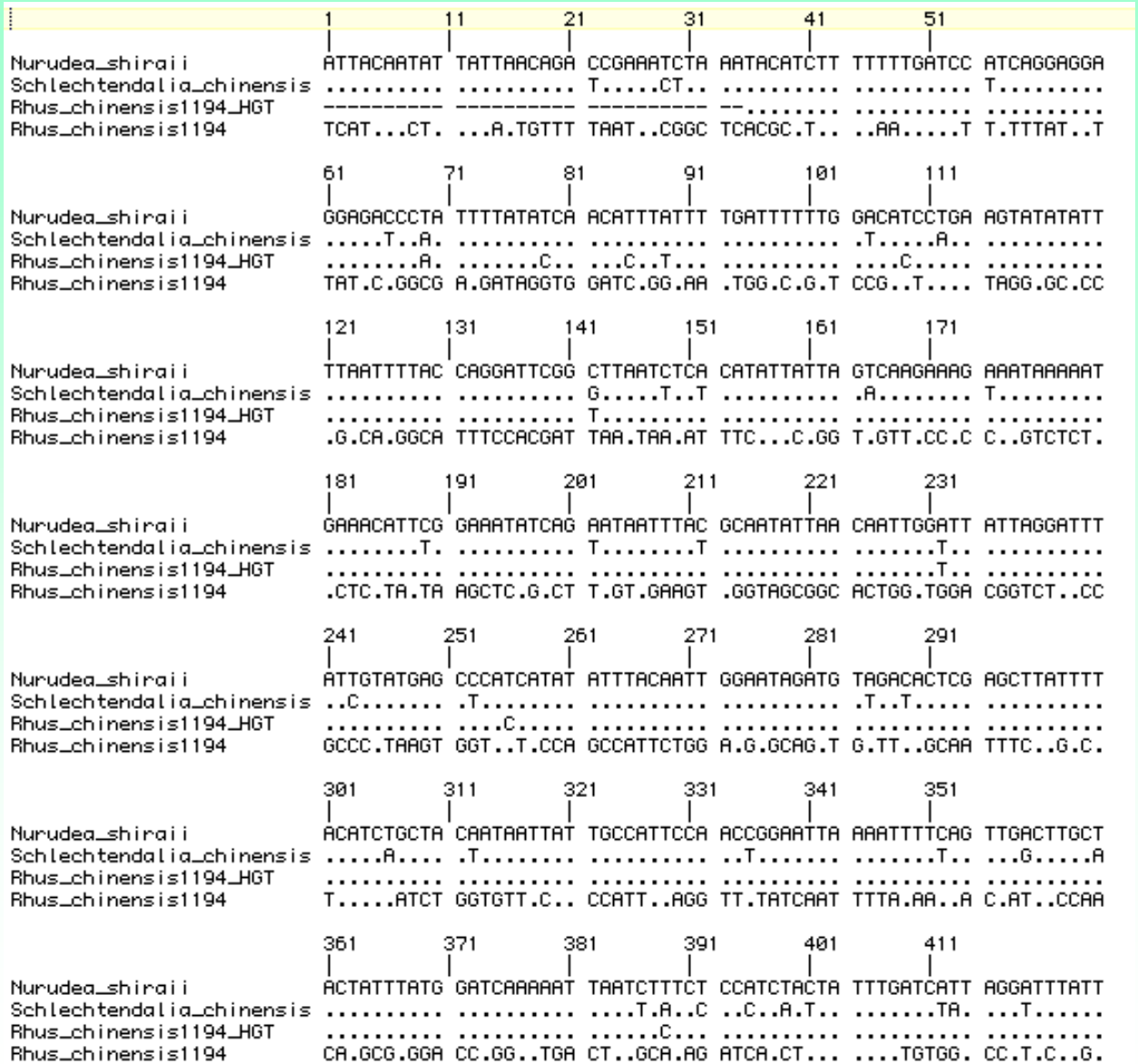
- **known as *Woo-pei-tsze* (五倍子) or *Chinese gall***
- **specially refer to the aphid group parasitizing on *Rhus* species to induce galls**
- **economically important in Asia as traditional medicines as well as sources of industrial tannin**
- **lifecycles with alternating sexual and parthenogenetic generations**
- **unique in alternating hosts with *Rhus* species as primary (summer) and mosses as secondary (winter) hosts**
- **Aphid-*Rhus* association: at least 48 million years according to fossil and biogeographic evidence (Moran, 1989)**

Diversity of parasitic-host relationships

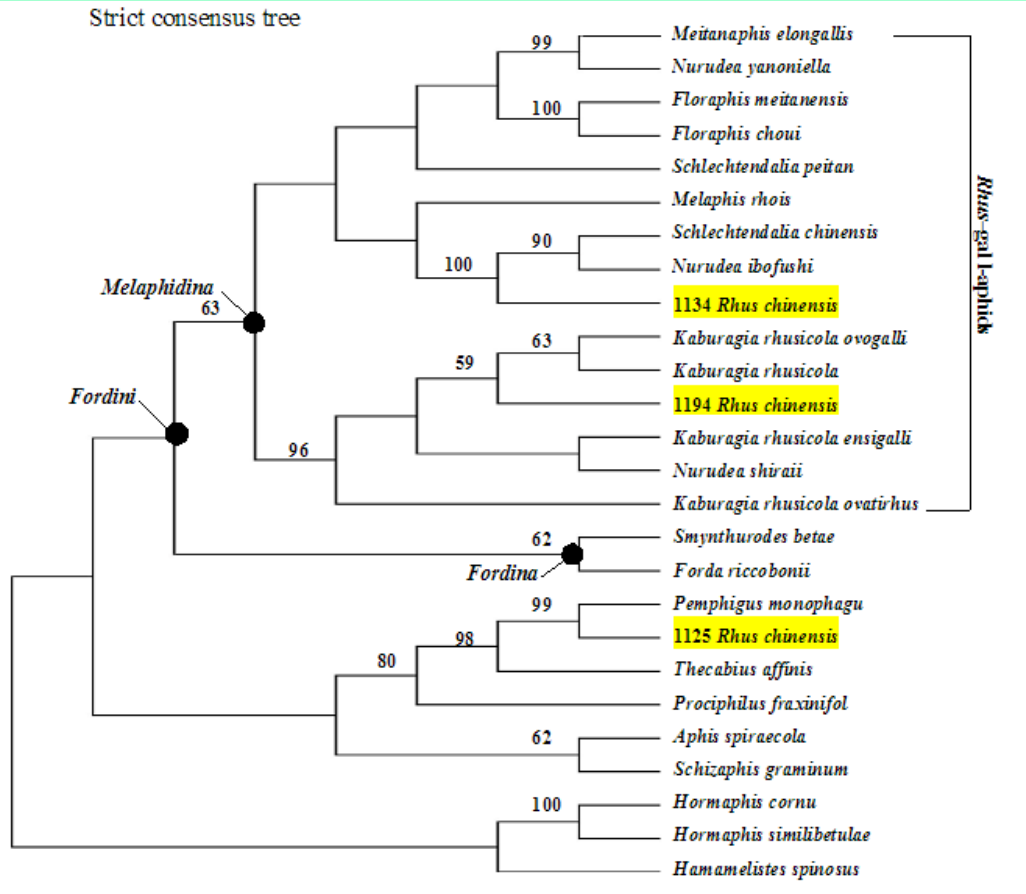
Rhus gall aphid		No of secondary hosts	Primary host
Schlechtendalia	*S. chinensis*	13	*Rhus chinensis*
	S. peitan	5	
Nurudea	*N. ibofushi*	NA	
	N. shiraii	8	
	N. yanoniella	NA	
Floraphis	*F. meitanensis*	1	*Rhus punjabensis*
	F. choui	NA	*Rhus potaninii*
Kaburagia	*K. rhusicola*	1	
	K. r. ovatirhusicola	2	
	K. r. ovogallis	12	*Rhus punjabensis*
	K. r. ensigallis	14	
Meitanaphis	*M. elongallis*	4	
	M. flavogallis	NA	
	M. microgallis	NA	*Rhus potaninii*
Melaphis	*M. rhois*	1	*Rhus typhina* *Rhus glabra*

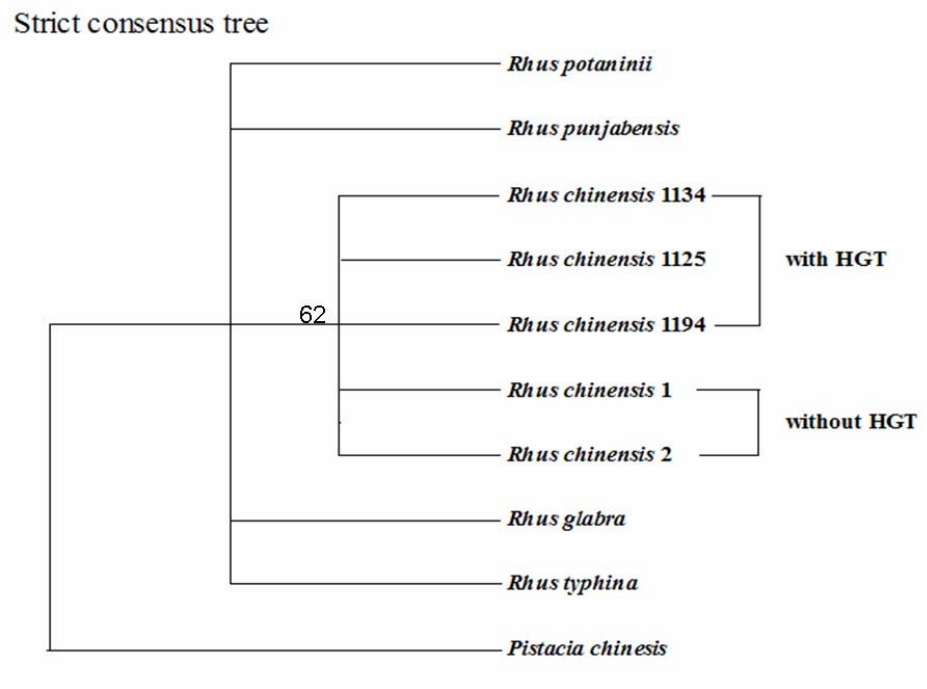
From Zhang(1999) and Eastop and Hille Ris Lambers (1976)

Evidence


1 11 21 31 41 51
Nurudea_shiraii ATTACAATAT TATTAACAGA CCGAAATCTA AATACATCTT TTTTTGATCC ATCAGGAGGA
Schlechtendalia_chinensis .... T.....CT.. .... T.....
Rhus_chinensis1194_HGT ---------- ---------- ---------- --..... ....
Rhus_chinensis1194 TCAT...CT. ...A.TGTTT TAAT..CGGC TCACGC.T.. ..AA.....T T.TTTAT..T
61 71 81 91 101 111
Nurudea_shiraii GGAGACCCTA TTTTATATCA ACATTTATTT TGATTTTTTG GACATCCTGA AGTATATATT
Schlechtendalia_chinensis .....T..A. .... .T.....A.. .....
Rhus_chinensis1194_HGT .....A. .....C.. ...C..T... ....C..... .....
Rhus_chinensis1194 TAT.C.GGCG A.GATAGGTG GATC.GG.AA .TGG.C.G.T CCG..T.... TAGG.GC.CC
121 131 141 151 161 171
Nurudea_shiraii TTAATTTTAC CAGGATTCGG CTTAATCTCA CATATTATTA GTCAAGAAAG AAATAAAAAT
Schlechtendalia_chinensis .... G.....T..T ..... .A..... T.....
Rhus_chinensis1194_HGT .... T..... ....
Rhus_chinensis1194 .G.CA.GGCA TTTCCACGAT TAA.TAA.AT TTC...C.GG T.GTT.CC.C C..GTCTCT.
181 191 201 211 221 231
Nurudea_shiraii GAAACATTCG GAAATATCAG AATAATTTAC GCAATATTAA CAATTGGATT ATTAGGATTT
Schlechtendalia_chinensis .....T. ..... T.....T ....T.. .....
Rhus_chinensis1194_HGT ....T.. .....
Rhus_chinensis1194 .CTC.TA.TA AGCTC.G.CT T.GT.GAAGT .GGTAGCGGC ACTGG.TGGA CGGTCT..CC
241 251 261 271 281 291
Nurudea_shiraii ATTGTATGAG CCCATCATAT ATTTACAATT GGAATAGATG TAGACACTCG AGCTTATTTT
Schlechtendalia_chinensis ..C..... .T..... .... .T..T..... .....
Rhus_chinensis1194_HGT ....C..... ....
Rhus_chinensis1194 GCCC.TAAGT GGT..T.CCA GCCATTCTGG A.G.GCAG.T G.TT..GCAA TTTC..G.C.
301 311 321 331 341 351
Nurudea_shiraii ACATCTGCTA CAATAATTAT TGCCATTCCA ACCGGAATTA AAATTTTCAG TTGACTTGCT
Schlechtendalia_chinensis .....A.... .T..... ....T..... .....T.. ...G.....A
Rhus_chinensis1194_HGT ....
Rhus_chinensis1194 T.....ATCT GGTGTT.C.. CCATT..AGG TT.TATCAAT TTTA.AA..A C.AT..CCAA
361 371 381 391 401 411
Nurudea_shiraii ACTATTTATG GATCAAAAAT TAATCTTTCT CCATCTACTA TTTGATCATT AGGATTTATT
Schlechtendalia_chinensis ....T.A..C ..C..A.T.. .....TA. ...T.....
Rhus_chinensis1194_HGT ....C... ....
Rhus_chinensis1194 CA.GCG.GGA CC.GG..TGA CT..GCA.AG ATCA.CT... ....TGTGG. CC.T.C..G.

Phylogenetic analyses





- **"Insertion" of one copy in *Rhus* taxa into the aphids tree**
- **Another copy is still in *Rhus* group**

Questions and Hypothesis

Samples:

Year 1:

- **Seedlings：200 individuals**
- **Adult Leaves：55 individuals**
- **5％ frequency for HGT**

Year 2:

- **Seeds：370 individuals**
- **Seedlings： 180 individuals**

Year 3:

- **Repeated samples： 200 individuals**

- **Gene transfer or (mt) Genome transfer?**

Detecting HGT with next generation sequencing technology

Horizontal Gene Transfer by the Parasitic Plant *Striga hermonthica*

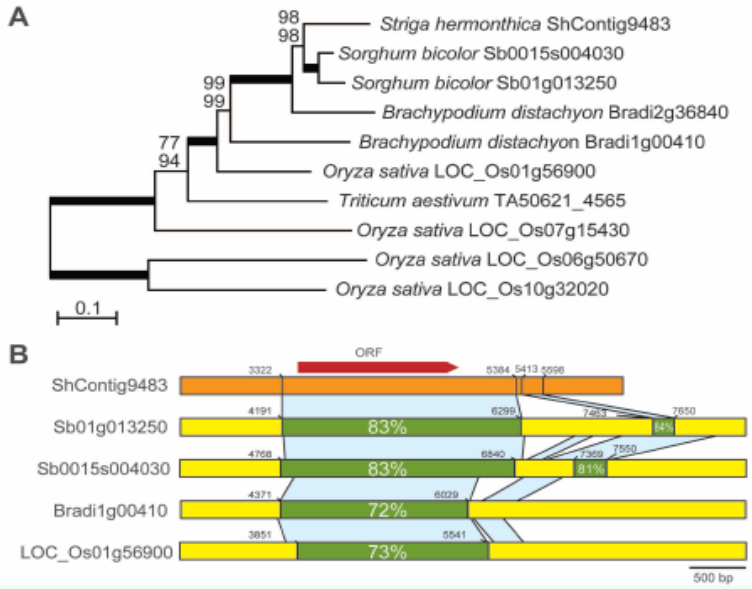
Satoko Yoshida,[1] Shinichiro Maruyama,[2] Hisayoshi Nozaki,[2] Ken Shirasu[1]*

Horizontal gene transfer (HGT) plays an important role in genome evolution (*1*). In plants, the majority of reported cases of HGT have been limited to exchanges between plants and microbes, mitochondrial transfer, or the translocation of mobile elements among related species (*1*). Parasitic plants are known to be vectors of mitochondrial HGT, but it has been unclear whether they also mediate nuclear HGT (*1*, *2*).

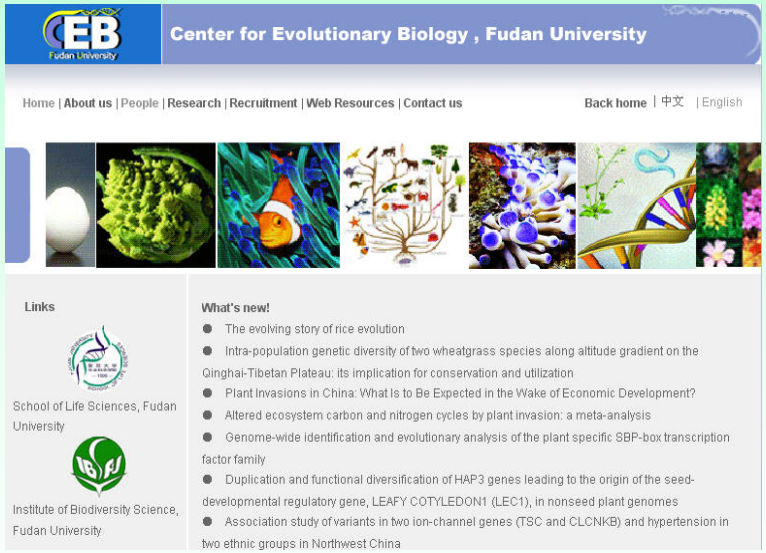
Striga hermonthica (Del.) Benth. is a devastating parasitic plant that infests members of the grass family (Poaceae), including major crops such as sorghum (*Sorghum bicolor*) and rice (*Oryza sativa*). *S. hermonthica* belongs to the eudicot Orobanchaceae family of the order Lamiales (fig. S1) (*3*) and only infects monocot plants. Thus, we reasoned that we may be able to detect nuclear HGT, if it occurs, by identifying monocot-specific genes in the *S. hermonthica* genome. From a large-scale expressed sequence tag analysis of *S. hermonthica* (*4*), we found one gene, designated *ShContig9483*, which shows high similarity to genes in sorghum and rice but has no homologs in eudicots (fig. S2A and table S1) (*5*). Southern blot analysis revealed *ShContig9483* cross-hybridization signals from sorghum, and rice to a lesser extent, reflecting its lower similarity, whereas no signals were detected from other closely related plants in Orobanchaceae nor from any nonparasitic eudicots (fig. S2B). This indicates that *ShContig9483* most likely originated in the monocots before its transfer to *S. hermonthica*.

ShContig9483 encodes a 448–amino acid protein with unknown function. Phylogenetic analysis of *ShContig9483* and related protein-coding sequences clusters *S. hermonthica* with sorghum (Fig. 1A). This tree conflicts with the phylogenetic position of *Striga* (*3*), suggesting that *S. hermonthica* acquired *ShContig9483* from sorghum or a related grass species. The *S. hermonthica* genomic region containing *ShContig9483* resides near a nuclear gene encoding a putative cis-prenyltransferase (fig. S3A). In contrast to *ShContig9483*, this putative cis-prenyltransferase gene from *S. hermonthica* clusters with genes from other eudicot species (fig. S3B).

The genomic sequences of the sorghum homologs show similarities to the *ShContig9483* locus from about 150 base pairs (bp) up- and 800 bp downstream of the putative open reading frame (ORF), except for several insertions and deletions (Fig. 1B and fig. S4). The similarity to homologs from *Brachypodium* and rice covers a shorter region with lower identities (Fig. 1B). The high conservation of sequence between *S. hermonthica* and sorghum outside the ORF suggest that transfer was a relatively recent event. Indeed, a sequence highly similar to *ShContig9483* was found in *S. gesnerioides* but not in *Orobanche minor*, which is from a closely related genus (figs. S2B and S5). These data suggest that incorporation of the *ShContig9483* fragment occurred before speciation of *S. hermonthica* and *S. gesnerioides* but after differentiation of the genera *Striga* and *Orobanche*. Parasitic plants form an invasive organ called a haustorium, which interconnects their vasculature with that of their hosts to allow transfer of nutrients, water, and even mRNAs (*2*). Thus, one possibility is that *ShContig9483* was originally captured by *S. hermonthica* as mRNA or cDNA. Interestingly, we found 13 consecutive adenine (A) nucleotides at the 3′ end of the *ShContig9483* genomic region (fig. S4). This sequence may be indicative of a poly-A tail added to a transcriptional unit of the originally transferred gene. In any case, our comparative genomics analysis of a eudicot parasite and its monocot hosts presents a clear case for nuclear HGT. Because the method used in this study is limited to the



Yoshida et al. 2010, *Science* April 15



Welcome to Institute of Biodiversity Science and Geobiology, Tibet University in Lhasa & Center for Evolutionary Biology, Fudan University in Shanghai !