Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Katedra botaniky

Oddělení vývojové genetiky rostlin

Biofyzikální ústav AV ČR, v.v.i.

Doktorská disertační práce

HORIZONTÁLNÍ GENOVÝ PŘENOS Z BAKTÉRIÍ RODU *RALSTONIA* DO GENOMU ROSTLIN

Martina Talianová

Brno 2011

Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Katedra botaniky

a

Biofyzikální ústav AV ČR, v.v.i.

Horizontální genový přenos z baktérií rodu *Ralstonia* do genomu rostlin

Doktorská disertační práce

Doktorský studijní program

P1527 Biologie - Botanika

Školitel RNDr. Jitka Žlůvová, PhD.

Brno 2011 RNDr. Martina Talianová

Bibliografická identifikace

Jméno a příjmení autorky: RNDr. Martina Talianová

Název disertační práce: Horizontální genový přenos z baktérií rodu *Ralstonia* do genomu rostlin.

Název disertační práce v anglickém jazyce: Horizontal gene transfer from the bacteria of the genus *Ralstonia* into plant genomes.

Studijní program: Biologie - č. P1527

Studijní obor: Botanika – č. 1507V004

Školitel: RNDr. Jitka Žlůvová, PhD

Rok obhajoby: 2011

Klíčova slova:

horizontální genový přenos, molekulární fylogenetika, transformace, Ralstonia, antimikrobiální protein, fúze proteinových domén, vznik nových genů, síťová evoluce

Key words:

horizontal gene transfer, molecular phylogenetics, transformation, Ralstonia, antimicrobial protein, interkingdom gene fusion, novel genes, reticulate evolution

Prohlašuji, že jsem disertační práci vypracovala samostatně a použila jen uvedených pramenů a literatury.

V Brně dne …….

Poděkování

Ďakujem svojim školiteľom Dr. Jitke Žlůvovej a Dr. Bohuslavovi Janouškovi za trpezlivé odborné vedenie mojej dizertačnej práce, ďalej pánovi profesorovi Borisovi Vyskotovi za podporu nielen v obdobiach zdaru, ale aj vo chvíľach vedeckej frustrácie. Tiež ďakujem všetkým svojím kolegom za vytvorenie skvelého kolektívu a pracovnej pohody. A v neposlednom rade ďakujem svojmu priateľovi Michalovi, ktorý so mnou trpezlivo prečkal niekedy zdanlivo nekonečné čakanie na titul PhD, a svojmu ocinovi, ktorý celé moje štúdium prežíval akoby bolo jeho vlastné. Veľká vďaka za podporu patrí celej mojej úžasnej rodine.

Obsah

1. Úvod a cíle práce

Horizontální genový přenos (HGP) je v posledních desetiletích jedním z nejvíce diskutovaných témat evoluční biologie, protože do znační míry změnil náš pohled na evoluční historii genomů. HGP sehrává klíčovou úlohu v evoluci bakteriálních genomů - významně se podílí na schopnosti adaptace a speciace (Doolittle 1999; Keeling a Palmer 2008). Do jaké míry HGP tvaruje genomy moderních eukaryot, zůstává zatím neznámé. Názory ohledně dynamiky a důležitosti HGP pro eukaryotické organismy jsou značně kontroverzní. Nicméně, HGP byl v podstatě klíčovým procesem při vzniku eukaryot, kdy došlo k pohlcení bakterií, ze kterých vznikli plastidy a mitochondrie. Následně se jejich genetický materiál integroval v rámci hostitelského genomu s postupným (a u rostlin pořád přetrvávajícím) putováním DNA z organel do jádra (Keeling and Palmer 2008). Když však odhlédneme od přenosu DNA z organel, jádra dnešních eukaryot se zdají být poměrně odolná vůči přenosu cizorodé DNA. Příčinou je zjevně složitost a důmyslnost přirozených bariér, které zabraňují nekontrolovatelnému příjmu cizí DNA, a také různorodé mechanismy oddělení zárodečné linie. Výjimku tvoří poměrně častý HGP u jednobuněčných eukaryot, který je často spojený s fagotrofií, endosymbiózou, či s parazitickým životním stylem (Andersson, 2005; Keeling a Palmer 2008; Keeling 2009). Existuje však i mnoho příkladů HGP u mnohobuněčných eukaryot, jako například velké množství genů z různých donorů v genomu bdeloidních rotifer (Gladyshev et al. 2009), nebo přenos genů z vnitrobuněčné bakterie *Wolbachia* do genomu mnoha hostitelů ze skupiny členovců (Hotopp et al. 2007). Rostoucí počet takových případů ukazuje, že překážky přenosu cizorodé DNA mohou být překonány i u složitějších organismů. S přibývajícím množstvím sekvenčních dat a se zvyšující se citlivosti detekčních metod jsou objevovány další případy HGP zahrnujících mnohobuněčná eukaryota. Mnohé z těchto případů byly popsány u rostlin (Bock, 2010). Rostliny jsou opravdu velice dobrými kandidáty pro HGP - ve srovnání se zvířaty je jejich zárodeční linie více vystavená vnějším vlivům. Také genomy rostlin jsou mnohem plastičtější než genomy zvířat (Kejnovsky et al. 2009). Přehled problematiky horizontálního genového přenosu u rostlin je podrobně shrnutý v příloze I.

Cílem této práce byla analýza sekvencí, které byly horizontálně přeneseny z bakterií příbuzných druhu *Ralstonia solanacearum* do genomu rostlinných druhů. Jednalo se o tyto dva tematické celky:

- charakterizace a evoluční dynamika sekvence MK14 u druhů rodu *Silene* (Caryophyllaceae)

- identifikace dalších případů HGP mezi druhem *R. solanacearum* a rostlinami

2. Přehled získaných výsledků

2.1. Přehled metod molekulární fylogenetiky.

(Talianová 2007, příloha II)

Většina metod určených k detekci HGP spadá do 4 kategorií: (1) odlišné složení bazí v DNA, (2) neobvyklá fylogenetická distribuce, (3) neobvyklá sekvenční podobnost (t.j. větší podobnost s genem ze vzdáleně příbuzného druhu), a (4) inkongruentní fylogenetické stromy (e.g. rozporuplné genealogie genů) (Hervé a Douady 2003). Různé přístupy vedou k podezření, že došlo k HGP, nicméně detekce je nedostatečná, pokud není doprovázená řádnou fylogenetickou analýzou. Zlatým standardem pro identifikaci HGP je fylogenetická neshoda, ke které dochází tehdy, když je výrazný konflikt mezi fylogenezí genů a fylogenezí druhů. Existuje však několik matoucích faktorů, které je potřeba brát v úvahu. Abychom mohli spolehlivě usuzovat na fylogenetickou neshodu, je nutný důkladný taxonomický výběr pro daný gen, což je i v dnešní době často problém. V některých případech je těžké rozlišit mezi dávnou genovou duplikací a rozdílnou ztrátou genů. V jiných případech mohou být neobvyklé fylogenetické stromy zavádějící, protože rozdílné rychlosti evoluce mohou způsobovat artefakty ve fylogenetické rekonstrukci, a tím pádem vést k chybným stromům, které jsou silně statisticky podpořeny. Nicméně však značně roste množství případů, ve kterých dobře statisticky podpořené fylogeneze se širokým taxonomickým výběrem ukazují jasný konflikt s evolučními vztahy mezi organismy.

Tato práce představuje přehled fylogenetických metod, které se běžně používají k rekonstrukci evolučních vztahů. Jedná se o metody vhodné jak k analýze morfologických, tak především sekvenčních dat. U každé metody je uveden princip, na základě kterého metoda pracuje, a také je uvedeno, pro jaká data je či není metoda vhodná. Vzhledem k tomu, že oblast molekulární fylogenetiky je značně široká a metodický pokrok velice rychlý, slouží tato práce k základnímu přehledu pro ty, kteří s fylogenetickou analýzou teprve začínají.

2.2. Identifikace a charakterizace bakteriální sekvence v genomu několika druhů rostlinného rodu *Silene*

(Talianová et al. 2011, příloha III)

Rod *Silene* (Caryophyllaceae) je používán jako model pro studium různých evolučních a ekologických aspektů. Genomické zdroje pro některé druhy začali být dostupné pouze nedávno, což značně stimulovalo molekulárně-genetické studie u tohoto rodu. Mikrodisekce chromozomů a konstrukce genomických knihoven se běžně používá ke studiu genomů a ke hledání chromozomově-specifických markerů (e.g. Hobza a Vyskot, 2007). Mikrodisekce chromozomů byla také využita ke studiu struktury pohlavních chromozomů X a Y u *Silene latifolia* (knotovka bílá) (Delichère et al. 1999, Matsunaga et al. 1999, Hobza et al. 2004, Hobza et al. 2007). Touhle cestou bylo nalezeno několik zajímavých markerů pocházejících z Y-chromozom-specifické genomické knihovny *S. latifolia*, které byly podrobně zkoumány (Hobza et al. 2006, Kejnovský et al. 2006).

Naším cílem bylo charakterizovat sekvenci nazvanou MK14, která pochází z genomické plazmidové knihovny odvozené technikou Degenerate Oligonucleotid Primer (DOP)-PCR na Y-chromozomu *S. latifolia*. Porovnání sekvence MK14 s veřejnou databází sekvencí NCBI ukázalo silnou podobnost s částmi dvou sousedících genů bakteriálních druhů rodu *Ralstonia* (gram-negativní Betaproteobakterie, *Burkholderiaceae*). Porovnání se sekvencemi z rostlinné říše neukázalo žádnou statisticky významnou podobnost. Abychom vyloučili bakteriální kontaminaci, pomocí primerů navržených podle sekvence MK14 jsme PCR-amplifikovali příslušnou oblast jak z listů *S. latifolia, S. diclinis* a *S. dioica* pěstovaných ve skleníku, také na axenických kulturách odvozených z povrchově sterilizovaných semen *S. latifolia* a *S. vulgaris*. Tyto kultury rostly na médiu s obsahem antibiotika. Bakteriologické testy vyloučily přítomnost bakterií rodu *Ralstonia* v semenech *S. latifolia*. Pomocí fylogenetických analýz bylo potvrzeno, že sekvence MK14 ze *S. latifolia* je s vysokou statistickou významností evolučně příbuzná bakteriálním sekvencím druhů rodu *Ralstonia*. Southern blot hybridizace na genomické DNA samečků a samic dvou různých populací *S. latifolia* pomocí specifické sondy odvozené z bakteriální sekvence MK14 potvrdila přítomnost bakteriální inzerce v genomu *S. latifolia*. Zjistilo se, že tato inzerce není specifická pro Y-chromozom, a také byly prokázany délkové polymorfizmy mezi oběma populacemi.

Získané výsledky ukazují, že vzhledem k přítomnosti bakteriální inzerce u blízce příbuzných druhů *Silene*, tato inzerce byla s největší pravděpodobností přítomna u předka těchto druhů. Inzerce zřejmě pochází z bakteriálního předka druhů rodu *Ralstonia*. Současné druhy tohoto bakteriálního rodu jsou asociovány s různými ekologickými nikami (voda, půda, rhizosféra rostlin). Některé druhy jsou fytopatogenní anebo jsou asociovány s infekcemi u lidí se sníženou obranyschopností (Stelzmueller et al. 2006). Mechanismem odpovědným za asexuální přenos genetické informace mezi vzdáleně příbuznými organismy je horizontální genový přenos, a je možné předpokládat několik způsobů, jak se bakteriální DNA mohla začlenit do genomu druhů *Silene*. Nejpravděpodobnější možností je spontánní transformace během infekce (v případě, že bakteriální předek byl patogenní), nebo spontánní transformace během symbiózy. Oba způsoby přenosu byly pozorovány během experimentální transformace rostlin patogenními bakteriálními druhy *Agrobacterium tumefaciens, A. rhizogenes*, nebo symbiotickými druhy *Sinorhizobium meliloti*, *Mesorhizobium loti*, *Rhizobium* sp. (Broothaerts et al. 2005). Je obtížné určit přesný mechanismus HGP, protože se zřejmě jedná o dávnou událost. Je však zajímavé, že druhy rodu *Ralstonia* mají zvlášní druh mobilní DNA, katabolický transpozon pomocí kterého degradují bifenylové sloučeniny (Toussaint et al. 2003, Ryan et al. 2009). Oblast tohoto mobilního úseku obsahuje geny, které jsou asociované s patogenitou, a je známo, že tyto geny se podílejí na přenášení substrátových molekul do cílových buněk, včetně horizotálního přenosu DNA do jiných bakterií a eukaryot (Backert a Meyer 2006). Vzhledem k tomu, že bakterie jsou schopné si velice lehce vyměňovat geny, je možné, že kromě druhů *Agrobacterium* existují i jiné druhy schopné přenosu své DNA do eukaryotických jader. Exprese bakteriální DNA u *S. latifolia* nebyla prokázána, což může být způsobeno regulací exprese mechanismy rostlinného hostitele, nebo degenerací sekvence v případě, že hostiteli neposkytuje žádnou adaptivní výhodu.

2.3. Spojování proteinových domén pocházejících z různých říší v případě antimikrobiálního proteinu u lilku brambory (*Solanum tuberosum*).

(Talianova et al., příloha IV)

Značný pokrok v získávaní nových genomických dat z různých organismů výrazně posunul porozumění evoluce genů s novými funkcemi. Je známo několik mechanismů, které samostatně, nebo v kombinaci vytvářejí nové geny (Long et al. 2003). Jedním z nich je horizontální genový přenos (HGP), na který jsme se zaměřili v naší práci. Mnoho faktorů znesnadňuje detekci HGP, jako například rozdílná frekvence nukleotidových substitucí (Mirkin et al. 2003), ztráta genů specifická pro některou evoluční linii (Huang and Gogarten 2006), anebo ztráta fylogenetického signálu během evoluce. Navzdory tomu se několik studií pokusilo odhadnout rozsah HGP u eukaryot, a zjistit tak, do jaké míry je HGP pro eukaryotické organismy důležitý (e.g. Andersson 2005). Většina těchto studií zkoumala přítomnost celých genů u různých organismů, i když někteří autoři namítají, že tento přístup je příliš povrchní, aby bylo možné s větší mírou přesnosti identifikovat případy přenosu (Wolf et al. 2000, Choi and Kim 2007, Chan et al. 2009). Chan et al. (2009) ve studii prokaryotických genomů předpokládají, že horizontálně přenesené a rekombinované úseky DNA mohou kódovat strukturně neporušené proteinové domény, které mohou sloužit jako jednotky HGP. Autoři ukázali, že v rámci genomu může HGP doprovázen homologní rekombinaci a přestavět i funkčně velice konzervativní oblasti a vytvořit tak geny s mozaikovým původem.

Prohledáváním veřejné databáze NCBI za účelem identifikace dalších potenciálních případů HGP zahrnujících bakterie rodu *Ralstonia* se nám podařilo najít sekvenci antibakteriálního proteinu AP1 (a jemu odpovídajícího úseku DNA) izolovaného ze vzácné odrůdy lilku bramboru (*Solanum tuberosum*) (Feng et al. 2003). Protein AP1 poskytuje rostlinám rezistenci vůči několika kmenům bakterie *Ralsonia solanacearum* (jenž je závažným patogenem hospodářských plodin) a rezistenci vůči patogenním houbám *Rhizoctonia solani* a *Alternaria solani*. Feng et al. (2003) si všimli, že AP1 se skládá ze dvou odlišných proteinových domén. Sekvence na C-konci vykazuje podobnost k ATP-vázající doméně proteinu UspA (Universal stress protein A) (UspA doména). ATP-vázající doména obsahuje konzervativní motiv, který je příbuzný k nukleotid-vážícím proteinům a k proteinům

signalizačních drah (Zarembisky et al. 1998). N-terminus AP1 nevykazoval statisticky významnou podobnost k rostlinným proteinům, ale silnou podobnost k bakteriální kyselé fosfatáze (fosfoesterázová doména). Naším cílem bylo přesně určit evoluční původ obou domén proteinu AP1. Výhodou naší práce ve srovnání s Feng et al. (2003) je obrovské množství sekvenčních dat dostupné v databázích. Získáním velikého množství aminokyselinových sekvencí z nejrůznějších taxonomických skupin, které odpovídaly jednotlivým doménám proteinu AP1, srovnáním těchto sekvencí do souboru dat, a jejich následnou podrobnou fylogenetickou analýzou jsme ukázali, že obě proteinové domény, UspA doména a fosfoesterázová doména, mají jasně odlišný evoluční původ. UspA doména je jednoznačně rostlinného původu, zatímco fosfoesterázová doména ukazuje evoluční příbuznost k bakteriím rodu *Ralstonia*. Předpokládáme, že přítomnost bakteriální fosfoesterázové domény v proteinu AP1 byla způsobená dávným HGP z předka příbuzného bakteriím rodu *Ralstonia* a následnou reorganizací pomocí rekombinace, která spojila obě domény ve funkční celek. K HGP mohlo dojít různými způsoby, nicméně vzhledem k tomu, že se jedná o dávnou událost, není možné určit přesný mechanismus. Je taktéž možné předpokládat i jiné alternativy pro vysokou podobnost fosfoesterázové domény k bakteriálním sekvencí, které vylučují HGP (např. ztráta této domény z rostlinných genomů, nebo duplikaci rostlinné domény s následnou divergencí). Tyto hypotézy se však ukázaly být nedostatečně pravděpodobné.

Podrobná fylogenetická analýza antibakteriálního proteinu AP1 u *S. tuberosum* ukázala, že horizontální genový přenos a kombinace proteinových domén může být důležitým zdrojem genů s novou funkcí, a tím může přispět ke schopnosti adaptace eukaryot. Předpokládáme však, že u vyšších eukaryot k podobným událostem dochází s nízkou frekvencí. Tento případ podporuje tvrzení, že při identifikaci HGP by měla být věnována větší pozornost evoluční historii menších funkčních jednotek, než jsou celé geny, t.j. například proteinovým doménám.

3. Literatura

Andersson JO. 2005. Lateral gene transfer in eukaryotes. Cell Mol Life Sci **62**: 1182-1197

Backert S, Meyer TF. 2006. Type IV secretion systems and their effectors in bacterial pathogenesis. Curr Opin Microbiol **9**: 207-217

Bock R. 2010. The give-and-take of DNA: horizontal gene transfer in plants. Trends Plant Sci **15**: 11-22

Broothaerts W, Mitchell HJ, Weir B, Kaines S, Smith LM, Yang W, Mayer JE, Roa-Rodriguez C, Jefferson RA. 2005. Gene transfer to plants by diverse species of bacteria. Nature **433**: 629-633

Chan CX, Darling AE, Beiko RG, Ragan MA. 2009. Are protein domains modules of lateral genetic transfer? PloS ONE **4**: e4524

Choi I, Kim S. 2007. Global extent of horizontal gene transfer. Proc Natl Acad Sci USA **104**: 4489-4494

Delichère C, Veuskens J, Hernould M, Barbacar N, Mouras A, Negrutiu I, Monéger F. 1999. Sl*Y1*, the first active gene cloned from a plant Y chromosome, encodes a WD-repeat protein. EMBO J **18**: 4169-4179

Doolittle WF. 1999. Phylogenetic classification and the universal tree. Science **284**: 2124-2129

Feng J, Yuan F, Gao Y, Liang C, Xu J, Zhang C, He L. 2003. A novel antimicrobial protein isolated from potato (*Solanum tuberosum*) shares homology with an acid phosphatase. Biochem J **376**: 481-487

Gladyshev EA, Meselson M, Arkhipova IR. 2008. Massive horizontal gene transfer in bdelloid rotifers. Science **320**: 1210-1213

Hervé P, Douady CJ. 2003. Horizontal gene transfer and phylogenetics. Curr Opin Microbiol **6**: 498-505

Hobza R, Lengerova M, Cernohorska H, Rubes J, Vyskot B. 2004. FAST-FISH with laser beam microdissected DOP-PCR probe distinguishes the sex chromosomes of *Silene latifolia*. - Chromosome Res **12**: 245-250

Hobza R, Hrusakova P, Safar J, Bartos J, Janousek B, Zluvova J, Michu E, Dolezel J, Vyskot B. 2006. MK17, a specific marker closely linked to the gynoecium suppression region on the Y chromosome in *Silene latifolia*. Theor Appl Genet **113**: 280-287

Hobza R, Vyskot B. 2007. Laser microdissection-based analysis of plant sex chromosomes. Methods Cell Biol **82**: 433-453

Hobza R, Kejnovsky E, Vyskot B, Widmer A. 2007. The role of chromosomal rearrangements in the evolution of *Silene latifolia* sex chromosomes. Mol Genet Genomics **278**: 633-638

Hotopp JCD, Clark ME, Oliveira DCSG, Foster JM, Fischer P, Torres MCM, Giebel JD, et al. 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **317**: 1753-1756

Huang J, Gogarten JP. 2006. Ancient horizontal gene transfer can benefit phylogenetic reconstruction. Trends Genet **22**: 361-366

Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet **9**: 605-618

Keeling PJ. 2009. Functional and ecological impacts of horizontal gene transfer in eukaryotes. Curr Opin Genet Dev **19**: 613-619

Kejnovsky E, Leitch IJ, Leitch AR. 2009. Contrasting evolutionary dynamics between angiosperm and mammalian genomes. Trends Ecol Evol **24**: 572-582

Long M, Betrán E, Thornton K, Wang W. 2003. The origin of new genes: glimpses from the young and old. Nat Rev Genet **4**: 865-875

Matsunaga S, Kawano S, Michimoto T, Higashiyama T, Nakao S, Sakai A, Kuroiwa T. 1999. Semi-automatic laser beam microdissection of the Y chromosome and analysis of Y chromosome DNA in a dioecious plant, *Silene latifolia*. Plant Cell Physiol **40**: 60-68

Mirkin BG, Fenner TI, Galperin MY, Koonin EV. 2003. Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. BMC Evol Biol **3**: 2

Ryan M, Pembroke JT, Adley C. 2009. Novel Tn4371-ICE like element in *Ralstonia pickettii* and genome mining for comparative elements. BMC Microbiol **9**: 242

Stelzmueller I, Biebl M, Wiesmayr S, Eller M, Hoeller E, Fille M, Weiss G, Lass-Floerl C, Bonatti H. 2006. *Ralstonia pickettii* - innocent bystander or a potential threat? Clin Microbiol Infect **12**: 99-101

Toussaint A, Merlin C, Monchy S, Benotmane MA, Leplae R, Mergeay M, Springael D. 2003. The biphenyl- and 4-chlorobiphenyl-catabolic transposon Tn4371, a member of a new family of genomic islands related to IncP and Ti plasmids. Appl Environ Microbiol **69**: 4837-4845

Wolf YI, Kondrashov AS, Koonin EV. 2000. Interkingdom gene fusions. Genome Biol **1**: RESEARCH0013

Zarembinski TI, Hung LW, Mueller-Dieckmann HJ, Kim KK, Yokota H, Kim R, Kim SH. 1998. Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. Proc Natl Acad Sci USA **95**: 15189-15193

4. Shrnutí

Evoluční historii rostlin poměrně často (ve srovnání s evolucí zvířat a hub) doprovázely "síťové" evoluční události. Takové evoluční sítě jsou tvořeny interspecifickou hybridizací, což je proces charakteristický pro rostlinnou říši. Také horizontální genový přenos (HGP) může být považován za příčinu síťovité evoluce, i když u eukaryot k němu dochází zřejmě v malé četnosti. Značný pokrok sekvenačních technologií, komparativní genomiky, a zvětšující se množství dat umožňuje identifikaci nových případů HGP. Tyto jsou vzhledem k nedostatku experimentálních dat identifikovány většinou náhodně, na základě jejich důsledků, t.j. dochází k neshodě v evolučních vztazích rekonstruovaných použitím fylogenetických metod.

Naše práce dál přispěla ke zvyšujícímu se množství případů HGP zahrnujícímu vyšší eukaryota, a zejména rostliny. Byly potvrzeny a popsány dva případy spontánního horizontálního genového přenosu, který zahrnují bakterie z rodu *Ralstonia* a dva různé rostlinné hostitele, druhy rodu *Silene*, a lilek brambor (*Solanum tuberosum*). Tyto případy naznačují, že u eukaryot jsou přirozené bariéry, které zabraňují příjmu cizorodé DNA, pravděpodobně překonávány ve větší míře, než se předpokládalo. Co se týče spontánní bakteriální transformace rostlin, doposud je známý pouze případ transformace druhů z rodu *Nicotiana* bakteriálním druhem *Agrobacterium rhizogenes*. Na základě našich výsledků jemožno tudíž předpokládat, že druhy rodu *Agrobacterium* nejsou jediné bakterie schopné přenášet svou genetickou informaci do rostlinných genomů. Nesmírně zajímavé je zjištění, že horizontální genový přenos bakteriální DNA v souhře s dalšími procesy, jako například rekombinace mezi genomickými oblastmi, dal vznik mozaikovému antimikrobiálnímu proteinu AP1 u lilku brambory, který evidentně poskytuje rezistenci vůči bakteriálnímu donorovi DNA, a tím hostiteli poskytuje značnou adaptivní výhodu. Tento případ dále ukazuje, že při snaze identifikovat HGP je třeba věnovat zvýšenou pozornost evoluční historii menších strukturních jednotek (např. proteinových domén), než jsou celé geny. Nevýhodou těchto studií je, že se jedná o dávné události, a tak většinou není možné přesně identifikovat okolnosti a mechanismy, které se podílely na začlenění cizorodé DNA do hostitelského genomu. Studium případů spontánního HGP však poskytuje nejenom zajímavá vodítka, ale především otevírá otázky ohledně nových cest k šíření informačních molekul, transgenů, organel, či patogenů, a také ohledně začleňování a expresní kontrole spontánních transgenů v rostlinných genomech.

Abstract

Evolutionary history of plants was relatively frequently (when compared with the evolution of fungi and animals) accompanied by reticulate evolutionary events. Such reticulations can be caused by interspecific hybridization - a process that is characteristic for plant kingdom. Horizontal gene transfer (HGT) can also be regarded as a cause of reticulate evolution, even though in eukaryotes it appears to be rather infrequent. Enormous progress in sequencing technologies, comparative genomics along with the increasing amount of sequence data enables to detect novel cases of HGT. These are mostly ancient events identified randomly based on their consequences, i.e. phylogenetic incongruence in evolutionary relationships.

Our work further contributes to the increasing number of HGT cases involving higher eukaryotes. We confirmed and characterized two cases of spontaneous gene transfer. These cases involve donor bacteria from the genus *Ralstonia* and two plant hosts species from the genus *Silene* and potato (*Solanum tuberosum*). This suggests that the natural barriers preventing uncontrolled uptake of foreign DNA in eukaryotes can be probably circumvented at higher frequency than previously though. To date, only one case of spontaneous bacterial transformation of plants is known involving species from the genus *Nicotiana* and bacterial species *Agrobacterium rhizogenes*. Our results thus indicate that species from the genus *Agrobacterium* are not the only bacteria able to naturally transform plants. An interesting findings is that HGT in concert with other processes, e.g. recombination between genomic regions, gave rise to mosaic antimicrobial protein AP1 in potato. This protein confers resistance to the putative donor of DNA, and provides the host with appreciable adaptive advantage. This case further shows that while attempting to identify HGT events, attention should also be paid to evolutionary history of smaller structural units (e.g. protein domains) rather than to whole genes only. A drawback of these studies is that they deal with ancient events, and it is thus difficult to exactly identify circumstances and mechanisms that took part in the foreign DNA incorporation to the host genome. Studying spontaneous HGT does not only provide us with interesting links, but also opens questions regarding novel routes of spread of informational molecules, transgenes, organelles, or pathogens. It also provides clues concerning integration and the control of expression of spontaneous transgenes in plant genomes.

5. Přílohy

Seznam přiložených publikací:

Příloha I

Talianova M., Janousek B. What can we learn from tobacco and other Solanaceae about horizontal DNA transfer? American Journal of Botany (v tisku)

Příloha II

Talianova M. 2007. Survey of molecular phylogenetics. Plant Soil and Environment **53**: 413-416

Příloha III

Talianová M., Žlůvová J., Hobza R., Vyskot B., Janoušek B. Identification and characterization of a bacteria-like sequence in the genome of some species from the plant genus *Silene*. Biologia Plantarum (v tisku).

Příloha IV

Talianova M., Vyskot B., Janousek B. Interkingdom protein domain shuffling: the case of an antimicrobial protein in potato (Solanum tuberosum). Rukopis předložen v časopise Plant Systematics and Evolution

Příloha V

Abstrakta z konferencí

Talianova M. (2006) Survey of molecular phylogenetics. 4. Metodické dny, Srní, p. 9

Talianova M., Zluvova J., Janousek B, Hobza R., Vyskot B. (2007) Horizontal gene transfer from the bacterial pathogen *Ralstonia solanacearum* to the plant genome of *Silene latifolia*. 11th Evolutionary Biology Meeting, Marseilles, Francie, p. 80

Talianova M., Zluvova J, Michu E, Forest A, Kamau E, Nicholas M, Charlesworth D, Marais G. (2008) Analysis of the efficacy of selection in *Silene* species with different breeding systems and its implication for the evolution of dioecy. Annual Meeting of the Society for Molecular Biology and Evolution, Barcelona, Španělsko

Talianova M. Horizontal gene transfer - new cases in plants. (2009) 9th International Congress on Plant Molecular Biology, St. Louis, Missouri, USA, p. 340

Talianová M., Janoušek B. (2010) Horizontal DNA transfer - what can we learn from plants? 8. Konference doktoradnů experimentální biologie rostlin. ČSEB, Praha, p. 22

PŘÍLOHA I

What can we learn from tobacco and other Solanaceae about horizontal DNA transfer?

Talianova M., Janousek B.

American Journal of Botany - přijato do tisku

WHAT CAN WE LEARN FROM TOBACCO AND OTHER SOLANACEAE ABOUT HORIZONTAL DNA TRANSFER?[1]

MARTINA TALIANOVA [2], BOHUSLAV JANOUSEK

Department of Plant Developmental Genetics, Institute of Biophysics AS CR,

Kralovopolska 135, 612 65, Brno, Czech Republic

[1]Manuscript received ; revision accepted .

The authors thank Prof. S. S. Renner, Prof. B. Vyskot and the two anonymous reviewers for constructive and very helpful comments and suggestions. This work was supported by Czech Science Foundation (grants no. 521/08/0932 a 204/09/H002) and Institutional Research Plans (AV0Z50040507 and AV0Z50040702).

[2]Author for correspondence (e-mail: talianka18@ibp.cz)

In eukaryotic organisms, horizontal gene transfer (HGT) is regarded as an important though infrequent source of reticulate evolution. Many confirmed instances of natural HGT involving multicellular eukaryotes come from flowering plants. This review intends to provide a synthesis of current knowledge regarding HGT in higher plants with an emphasis on tobacco and other species in the Solanaceae family because there are numerous detailed reports concerning natural HGT events, involving various donors, in this family. Moreover, in depth experimental studies utilizing transgenic tobacco are of great importance for understanding this process. Valuable insights are offered concerning the mechanisms of HGT, the adaptive role and regulation of natural transgenes, and new routes for gene trafficking. With an increasing amount of data on HGT, a synthetic view is beginning to emerge.

Key words: *Agrobacterium*; grafting; horizontal gene transfer; plasmodesmata; reticulate evolution; Solanaceae; tobacco; viruses

Horizontal gene transfer (HGT) has been one of the most debated topics in evolutionary biology during the last two decades, because it has substantially challenged our view of the evolutionary history of genomes. In bacteria, HGT occurs at a high frequency and plays a key role in the shaping of bacterial genomes throughout evolution. Historically, HGT played a crucial role in eukaryote origins, which involved the engulfment of bacteria that subsequently gave rise to plastids and mitochondrion, an integration of their genetic material within the host genome, along with subsequent and still ongoing migration of organellar DNA to the nucleus (Keeling and Palmer, 2008). Contemporary eukaryotic nuclei seem relatively resistant to the transfer of foreign DNA (if transfer of DNA from organelles is excluded). This is apparently caused by the complexity and sophistication of natural barriers that prevent uncontrolled uptake of nucleic acids and due to various mechanisms of germline sequestration. An exception is the frequent DNA uptake by microbial eukaryotes, which is often associated with a phagotrophic lifestyle, endosymbiosis, or parasitism (Andersson, 2005; Keeling and Palmer, 2008; Keeling, 2009). Nevertheless, there are numerous examples of HGT involving multicellular eukaryotes. Examples include the large number of genes from various donors in bdelloid rotifers (Gladyshev et al., 2008) or gene transfer from intracellular bacteria *Wolbachia* into several of its arthropod hosts (Hotopp et al., 2007). The growing number of such examples suggests that obstacles to the uptake of foreign DNA can be circumvented in rather complex organisms. Moreover, with an increasing amount of sequencing data, new instances of HGT involving multicellular eukaryotes are being discovered. Many of these have been reported in plants (recently reviewed by Bock, 2010). Indeed, plants are very good candidates for HGT; compared to animals, their germline is more exposed to external agents, and genomes of plants appear to be more plastic than those in animals (e.g., Kejnovsky et al., 2009).

In this short review, we discuss nuclear HGT in flowering plants with an emphasis on tobacco (*Nicotiana tabacum*), its close *Nicotiana* relatives and other Solanaceae. There are several reasons why we decided to focus our attention on the representatives of this family. Although the frequency of HGT in eukaryotes has been assumed to be very low, there are several reports of foreign genes from various donors (bacteria and viruses) being incorporated in the genome of solanaceous species. Moreover, in some species, foreign insertions have been experimentally investigated in detail, providing valuable clues about the mechanisms of HGT. So far, the detection of numerous natural nuclear HGT events within one taxonomic group is rather exceptional (in contrast to HGT of mitochondrial genes). The greater number of examples in solanaceous species may be largely due to the fact that many of these species are economically important and more attention has been focused on their investigation. In addition, in depth experimental studies related to various HGT issues have been performed on tobacco, which is also one of the best model species in transgenic research. We review the studies concerning cases of natural HGT involving various donors (bacteria and viruses) as well as experiments on HGT mechanisms using artificially transformed tobacco.

HGT from bacteria— The most well-known HGT involving bacteria and higher plants involves plant transformation by *Agrobacterium* species (*A. tumefaciens* and *A. rhizogenes*). While a detailed account of man-made transformation of plants is outside the scope of this review, it is worth briefly mentioning the procedure, which has become a widely used tool in molecular biology research and biotechnology (e.g., Tzfira and Citovsky, 2006). Moreover, most of the information about the molecular mechanism of DNA transfer comes from the studies on *Agrobacterium*-mediated plant transformation.

During transformation, the production of T-DNA and its transport into plant cells is promoted by a set of genes located on the bacterial chromosome and Ti (Ri) plasmid. Host proteins

primarily direct T-DNA transport within the cell, its import to the nucleus and integration in sites of double-stranded breaks by non-homologous recombination (e.g., Karami et al., 2009; Gelvin 2010). Tobacco, like many other Solanaceae species can be transformed both by *A. tumefaciens* and *A. rhizogenes* (e.g., Visser et al., 1989; de la Riva et al., 1998; Tzfira and Citovsky, 2006; Chen et al., 2007). Broothaerts et al. (2005) have shown that tobacco is also susceptible to transformation by other non-*Agrobacterium* species such as *Rhizobium sp.*, *Sinorhizobium meliloti*, *Mesorzhizobium loti* when these were supplied by the disarmed Ti-plasmid and a binary vector.

Almost three decades ago a natural transfer of DNA from *Agrobacterium rhizogenes* was reported to have occured during the evolution of the genus *Nicotiana* (White et al., 1985; Furner et al., 1986). It was most probably an ancient infection by an *A. rhizogenes*-like bacteria that left a footprint in at least 15 species (out of 42 species investigated) belonging to distinct sections of the genus *Nicotiana* as suggested by PCR analysis and Southern hybridization (Intrieri and Buiatti, 2001). Phylogenetic studies based on corresponding genes from different strains of *A. rhizogenes* and different *Nicotiana* species suggest that there have been at least two independent cases of horizontal gene transfer of T-DNA from *A. rhizogenes* to the genomes of *Nicotiana* species (i. e., natural transformation) in the history of the genus *Nicotiana* (Intrieri and Buiatti, 2001; Suzuki et al., 2002). This implies that the genus *Nicotiana* is probably more prone to horizontal gene transfer from *A. rhizogenes* to its genome in comparison with other angiosperms as no such cases were reported in other plant genera.

Cases of natural transformation by *A. rhizogenes* were studied in detail in *N. glauca* (Furner et al., 1986), where transferred genes comprise a region homologous to the pRi (root-inducing plasmid) T-DNA with several ORFs: *rolB*, *rolC*, ORF13 and ORF14 ordered as an incomplete inverted repeat. The region containing ORFs in *N. glauca* is called cT-DNA

(cellular T-DNA) and ORFs were correspondingly denoted as Ng*rolB*, Ng*rolC*, NgORF13 and NgORF14. Such homologous ORFs have been also found in wild-type *N. tabacum* (Meyer et al., 1995; Frundt et al., 1998). Artificially transformed plants (tobacco, carrot or morning glory) containing pRi of *A. rhizogenes* displayed an abnormal phenotype known as hairy root syndrome, exhibiting increased root mass, dwarfism due to reduced internode length, reduced apical dominance, wrinkled leaves or altered flower morphology (David et al., 1984; Tepfer 1984). It has been surmised that the insertion of pRi T-DNA might have contributed to speciation within the genus *Nicotiana*, since some of the transformation symptoms could confer an adaptive advantage to the species (e.g., increased root mass) (Suzuki et al., 2002). However, it has been demonstrated that inserted T-DNA genes have undergone rearrangement events (duplications, insertions, deletions) in each species since the divergence of *Nicotiana* species from the common ancestor that harboured the insertion (Intrieri and Buiatti, 2001; Suzuki et al., 2002) resulting in variable transcription patterns of these genes (Intrieri and Buiatti, 2001). In *N. glauca* for example, N*grol* genes are all transcribed to some extent, but the expression does not result in an abnormal phenotype (Aoki and Syono, 1999b; Aoki, 2004), thus there is no competitive advantage to be gained. The lack of phenotypic changes have been attributed to the truncation of the Ng*rolB* gene that contains a premature stop codon when compared to the original *rolB* gene of *A. rhizogenes* (Aoki and Syono, 1999b; Aoki, 2004).

Interestingly, in hybrids of *N. glauca* x *N. langsdorffii*, as well as in many other *Nicotiana* hybrids in which one parent carries *rol* genes (i.e., those originated by natural transformation) and the second parent lacks these genes, genetic tumors are spontaneously formed (e.g., Ahuja, 1962; Smith, 1988). The *rol* genes are thought to be involved in the regulation of growth factor synthesis and tumorogenesis (Aoki et al., 1994; Nagata et al., 1995; Udagawa, 2004). Indeed, Ng*ORF13*, Ng*ORF14*, Ng*rolB* and Ng*rolC* genes are expressed in genetic stem

tumors of the *N. glauca x N. langsdorffii* hybrid, while their transcripts were not detected in leaves (Aoki et al., 1994), which suggests that the expression of *rol* genes is strongly regulated. It is most probably a disturbance in control that results in tumor formation (e.g., Aoki et al., 1994; Ahuja, 1996; Udagawa et al., 2004). Environmental conditions are thought to play an important role in the formation of tumors and altered phenotype in interspecific hybrids. One of the suggested regulatory mechanisms is epigenetic control by DNA methylation as indicated by experiments with 5-azacytidine treatment of cell cultures originating from nontumorous *N. glauca x N. langsdorffii* hybrids where hypomethylation triggered tumor induction (Durante et al., 1989). Direct testing of the role of DNA methylation and other epigenetic mechanisms in the control of expression in genes horizontally transferred into *Nicotiana* genome from the genus *Agrobacterium* still remains to be performed.

***HGT from viruses*—** Even though bacteria are generally regarded as the most frequent donors of DNA to eukaryotes, footprints of viruses in the genomes of flowering plants also appear to be relatively frequent (Table 1). There are several reports of incorporation of viral DNA into the genomes of species from the genus *Nicotiana*. A well-studied case is the integration of geminiviral DNA into the genome of *Nicotiana tabacum* (inherited from its *N. tomentosiformis* paternal parent) and its three diploid relatives N. *kawakamii, N. tomentosa* and *N. tomentosiformis* (from the section *Tomentosae*) (Day et al., 1991; Bejarano et al., 1996; Ashby et al., 1997; Murad et al., 2004). Geminiviruses are insect-transmitted single-stranded DNA viruses that are able to infect a wide range of plants (Lazarowitz, 1992). Geminivirus-related DNA (GRD) in *Nicotiana* genomes is thought to arise through illegitimate integration of geminiviral DNA into the nuclear genome of an ancestor followed by amplification, divergence by deletions, and rearrangements. GRD's are arrayed as multiple

direct repeats and are composed of a degenerated geminiviral *rep* gene and a non-transcribed intergenic region carrying the origin of replication (*ori*). These repeats can be classified into two main families (denoted as GRD3 and GRD5), each of which contains blocks of several tens of direct tandem repeats (Murad et al., 2004). It is hypothesized that these families originated from two independent integration events into the ancestor of the *Tomentosae* species (Murad et al., 2002). Although individual members of both families are methylated and diverged, it has been speculated that they may have originally been involved in the host plant's resistance to viral infections (Murad et al., 2004).

Further studies have identified the presence of pararetroviral sequences, called endogenous pararetrovirus (EPRV) sequences, in the genomes of *Nicotiana* species as well as in several other members of *Solanaceae*. Pararetroviruses possess dsDNA and are considered retroelements due to their use of reverse transcription during replication. They usually, but not always, lack an integrase activity (Richert-Pöggeler and Shepherd, 1997). EPRVs are assumed to integrate via illegitimate recombination into the genome of their hosts where they can accumulate to high copy number (Harper et al., 2002). Jakowitsch et al. (1999) discovered pararetrovirus-like sequences integrated into the tobacco genome by illegitimate recombination. There are at least two families of EPRVs associated with polyploid tobacco, presumably inherited from its parental species. One family is supposed to have been integrated into *N. sylvestris* genome before polyploid formation (Jakowitsch et al., 1999), showing similar copy number, arrangement and DNA methylation pattern as in *N. tabacum* (Mette et al., 2002). The second family is represented by insertions integrated to *N. tomentosiformis* that show significantly increased abundance compared to *N. tabacum* due to either elimination from the tobacco genome or expansion in the *N. tomentosiformis* genome after polyploid formation (Gregor et al., 2004). Lockhart et al. (2000) identified an endogenous form of tobacco-vein-clearing-virus (TVCV) in *N. edwardsonii*. In this hexaploid

hybrid between *N. glutinosa* and *N. clevelandii*, activation of EPRVs occurs resulting in episomal infectious viral particles. However, in *N. glutinosa* (the paternal parent from which the EPRVs have been inherited; no copies are present in *N. clevelandii*) the copies are silent, showing no traces of virus infection. It is thought that endogenous EPRVs are kept under control by gene silencing mechanisms and may confer resistance to the virus (Matzke et al., 2004; Staginnus and Richert-Pöggeler, 2006).

Pararetrovirus-like sequences are also integrated in the genomes of other solanaceous species, such as petunia (Richert-Pöggeler and Shepherd, 1997), potato (*Solanum tuberosum*) (Hansen et al., 2005) and tomato (*Solanum lycopersicum*) (Staginnus et al., 2007). EPRVs integrated into plant genomes are often non-functional copies. In petunia, tomato and tobacco, EPRVs are often cytosine-methylated, suggesting transcriptional silencing. In some cases, EPRVs can be activated and can lead to infection symptoms (e.g., Richert-Pöggeler and Shepherd, 1997; Lockhart et al., 2000). Activation has often been observed in interspecific hybrids and could be triggered by stress conditions (e.g., changes in the light regime (Richert- Pöggeler and Shepherd, 1997; Lockhart et al., 2000)). Homology-dependent gene silencing has been hypothesized as a possible adaptive role for EPRVs as a mechanism of host defense against viruses (Jakowitsch et al., 1999; Mette et al., 2002; Staginnus et al., 2007).

Plant-to-plant exchange of genetic material— There is an increasing number of reports concerning natural HGT between higher plants. Most of the reported cases of HGT in flowering plants are either associated with parasitism and/or they involve mitochondrial sequences. Exchange of mitochondrial genes can occur in diverse directions involving various plant groups (angiosperms, gymnosperms, mosses, ferns), and most of the transfers are thought to be of recent origin (Bergthorsson et al., 2003; Won and Renner, 2003; Davis and Wurdack, 2004; Mower et al., 2004; Nickrent et al., 2004; Berghorsson et al., 2004;

Woloszynska et al., 2004; Schönenberger et al., 2005; Davis et al., 2005; Barkman et al., 2007) (Appendix S1, see Supplemental Data with the online version of this article). A unique case is the massive uptake of foreign mitochondrial genes from a wide range of plant donors by a basal angiosperm *Amborella trichopoda* (Bergthorsson et al., 2004). Shrubs of *A. trichopoda* live in diverse epiphytic associations that could serve as potential routes for DNA exchange other than parasitism. Interestingly, in contrast to the intensive flow of mitochondrial genes involving *A. trichopoda*, there is no report of chloroplast DNA transfers in this species (Goremykin et al., 2003; Rice and Palmer, 2006). This appears to be a general trend that holds for land plants (Rice and Palmer, 2006). To date, there has been no evidence for the between-species HGT of plastid genes despite the availability of numerous plant plastid genomes (Rice and Palmer, 2006; Keeling, 2010). This contrasting pattern has been attributed to the fact that unlike plastids, mitochondria are able to actively take up the DNA (Koulintchenko, 2003) and have been shown to often fuse with each other (e.g., Arimura et al., 2004). Interestingly, in spite of the abundance of mtDNA HGT in land plants, there are only negative reports concerning such transfer in tobacco, whose mitochondrial genome has been completely sequenced, and in other Solanaceae (e.g., the absence of a homing group I intron in the mitochondrial *cox1* gene in Solanaceae as shown by Sanchez-Puerta (2008)).

Further cases of plant-to-plant transfer involve transmission of mobile elements between flowering plants, thought to be mediated by some kind of a vector (Table 2). Recently, Yoshida (2010) reported on a natural HGT of a nuclear gene encoding a protein with unknown function from the parasitic plant *Striga hermonthica* (from eudicot Orobanchaceae family) to its monocot crop host *Sorghum bicolor*. These findings indicate that cell-to-cell contact established between plant species provides another possible route for gene flow through cell boundaries.

Valuable insights into gene transfer mediated by cell-to-cell contact in plants come from tissue grafting experiments. Tissue grafting, which shares some similarities to parasitic and symbiotic associations between plants, is widely used in plant breeding; however, similar associations also occur in nature (e.g., root-mediated grafting in trees (e.g., Jelínková et al., 2009)). Since Darwin's era, grafting has been supposed to give rise to "graft hybrids" (Darwin, 1883). The term "graft hybrids" was chosen because they arose from grafted plants and showed characters of both graft partners. In general, two different groups of "graft hybrids" can be distinguished according to their origin. The first case is so called "vegetative hybrids" that are claimed to be result of the "mentor grafting" (e.g., a seedling at cotyledonary phase is grafted onto an older stock; leaves of the scion except for two to three at the top are removed during the entire growth) (reviewed by Liu, 2006). They represent a controversial topic as their existence would mean that not only the phenotype but also genetic properties of scion can be influenced by stock. There is a suspicion that these results were often due to "ignorance or deliberate deception" and therefore the view of genetic community concerning "vegetative hybrids" has been skeptical (e.g., Redei, 1998). Although more recent efforts (spent in, e.g., eggplant, red pepper, petunia, tomato or tobacco) have generated some interesting and more credible observations (reviewed by Taller et al. 1998 and Liu 2006), a clear-cut interpretation of these observations is still not possible because no reliable molecular data are available. There is, however, experimental evidence that RNA molecules can travel long distances in a phloem sap and at least some of them are even accumulated in the shoot meristem (Ruiz-Medrano et al., 1999). The discovery of mRNA traffic in phloem, the role of RNAi (reviewed by Qui and Hanon, 2005), and findings in epigenetics (reviewed by Richards, 2006) indicate that some cases mimicking "vegetative hybridization" could occur. Occasional reverse transcription of the mRNA coming from the rootstock could of course even result in the heritable transfer of the acquired information.

The second group of "graft hybrids" is represented by periclinal chimeras that originate by regeneration from the boundary between the stock and the scion (reviewed, e.g., by Lee and Oda, 2003). Their origin can be stimulated by the removal of almost the whole scion. Until 2009, it was generally assumed that the cells in the chimera completely retain the genetic identity of the original "parents" (rootstock or scion) and that no transfer of genetic information occurs. However, Stegemann and Bock (2009) provided the first molecular support for nucleic acid transfer by performing grafting experiments with two transgenic tobacco lines. One line carried the kanamycin resistance gene and the yellow fluorescent protein reporter gene in the nucleus, while the second carried the spectinomycin resistance gene and the green fluorescent protein gene in its chloroplast genome. They demonstrated, that both reciprocal grafting procedures resulted in the occurrence of cells harboring both marker and reporter genes, and that these cells occurred in the graft junction area only. Transferred genes were shown to be stably inherited in subsequent generations of plants recovered from cells originating from the graft site, suggesting that changes can become heritable only via formation of lateral shoots from the graft junction. Stegemann and Bock (2009) further showed that the direction of gene transfer was solely from plastids to the nucleus, indicating that DNA can be transferred as large fragments or even as the entire chloroplast genome. While the mechanisms of DNA transfer between cells are largely unknown, plasmodesmata are speculated to be involved in the transfer (Stegemann and Bock, 2009).

Intracellular DNA transfer from chloroplasts to the nucleus— Intracellular gene transfer (IGT) from both mitochondria and chloroplasts is a widespread and ongoing process in flowering plants (e.g., Timmis et al., 2004; Leister, 2005; Noutsos, 2005; Richardson and Palmer, 2007; Bock and Timmis, 2008). Although this kind of gene transfer is not HGT *in*

sensu stricto, it is worth mentioning experimental studies focused on gene trafficking from chloroplasts to the nucleus in tobacco, as they provide us with important clues about mechanisms that can be extended to the non-organellar gene transfer. Stegemann et al. (2003) and Huang et al. (2003) designed a clever experimental system based on the transformation of tobacco chloroplast DNA with a specific construct bearing two antibiotic resistance genes (one gene with a chloroplast specific promoter and a second gene with a nuclear specific promoter). Leveraging the ability of tobacco plants to regenerate from a single transformed cell, the experimenters were able to create cycles of strong selection for antibiotic resistant plants. These experiments provided evidence that the invasion of chloroplast DNA to the nucleus occurs at an unexpectedly high rate at approximately 1 in 5,000,000 somatic cells (Stegemann et al., 2003), and 1 in 16,000 pollen grains (Huang et al., 2003; a similar order of magnitude was detected in Sheppard et al., 2008). Such elevated rates of transfer in male gametes are attributed to the same mechanisms that serve to prevent paternal inheritance of chloroplast; degradation of the plastid genome during male gametogenesis might serve as a source of DNA fragments available to enter the nucleus of male gametes (Nagata et al., 1999; Sheppard et al., 2008). Moreover, these results suggest that DNA travels predominantly as a bulk of chloroplast DNA (rather than mRNA or cDNA, as is often a case of gene transfers from mitochondria to nucleus (e.g., Nugent and Palmer, 1991; Adams et al., 2000)) from the plastid and recombines into a chromosome in the nucleus. Huang et al. (2003, 2004) used transplastomic tobacco plants to characterize *de novo* insertions of chloroplast DNA in the nucleus and revealed signs of insertions that resulted from double-strand break repair mediated by non-homologous recombination giving rise to loci with various degrees of complexity. Stegemann and Bock (2006) showed that successful activation of the marker gene in the nucleus is accompanied by several types of rearrangement such as deletions on short directly repeated sequences, deletions without homology at the breakpoints, point mutations,

and insertions related to double-strand breaks repair. These rearrangements are a consequence of capturing a nuclear gene promoter right upstream of the chloroplast gene (since promoter trapping is of key importance for enhancing transcriptional activity), often resulting in fused reading frames with upstream nuclear genes. Highly AT-rich non-coding regions downstream of chloroplast genes provide ready-to-use sites for RNA cleavage and polyadenylation, further contributing to the success of chloroplast-to-nucleus DNA transfer. However, the nuclear genome seems to effectively eliminate those sequences that do not become functionally activated (Huang et al., 2004; Stegemann and Bock, 2006). Deletions on short direct repeats and accumulation of point mutations are two mechanisms suggested to be involved in the rapid degeneration and elimination of transferred sequences in tobacco. Attention has so far been focused on the escape of DNA from organelles and their incorporation into the nuclear DNA as well as the fate of integrated sequences. The mechanisms that enable DNA fragments to enter the nucleus remain, however, to be experimentally elucidated.

***HGT involving fungi*—** Despite many differences between land plants and fungi, both groups are communally involved in tight ecological associations (e.g., mycorrhiza, endosymbiosis or commensalism, host-pathogen interactions, etc.) and thus ample opportunities for the exchange of genes are provided. Very close cell-to-cell contact is present in the system: mycorrhizal fungal symbiont-plant symbiont and parasitic fungus-plant host. There are indications that mycorrhizal fungi can serve as vectors of gene transfer between vascular plant species (e.g, between fern Botrychium virginianum (L.) and an unknown member of the parasitic order Santalales; Davis et al., 2005). For many cases of HGT, the mechanism of DNA transfer is not known (e. g., Bergthorson et al. 2003, Won and Renner 2003), and the role of fungi has been considered a viable possibility. The predisposition of fungi to work as putative vectors was stressed by Davis et al. (2005).

Indeed, mycorrhizas has been detected in 94% angiosperm plant families and in 85% of angiosperm species (Wang and Qiu, 2006). Moreover, it is known that mycorrhizal fungi are unspecific in their host selection and they therefore connect many distantly related plants in the same community (Smith and Read, 1997). However, the possibility, that fungi take part in HGT involving land plants, has not been largely explored to date. To our knowledge, the only systematic survey addressing this issue is the study of Richards et al. (2009). Their analysis indicates that the exchange of genetic information between plants (particularly the angiosperms) and fungi is rare, though ancient transfers occurred during the plant evolution and provided important adaptive phenotypes to both fungi and plants. Convincing reports of recent transfers between fungal and angiosperm lineages are, to our knowledge, lacking. However, it has been suggested that the group I intron in the mitochondrial *cox1* gene was recently horizontally transferred from fungi to some undetermined angiosperm and subsequently underwent extensive HGT to other lineages of flowering plants (Vaughn et al., 1995; Cho et al., 1998). It is questionable why there is such an extremely low rate of HGT between fungi and angiosperms being detected, given that fungi are not at all resistant to HGT involving various counterparts (e.g., Rosewich and Kistler, 2000; Jones et al., 2005). Are there specific barriers preventing an exchange, or does the lack of convincing cases point to insufficient sampling and a lack of resolution of analytical methods for HGT detection? It would be worthwhile to examine these findings in the light of cumulative genomic data.

DISCUSSION

The discovery of horizontal gene transfer significantly changed our view of the processes governing the evolution of organisms, both prokaryotic and eukaryotic. The frequency of HGT in eukaryotes has been a matter of discussions among evolutionary biologists, and the

increasing amount of confirmed cases of genes in eukaryotes acquired via HGT supports the idea that it is an ongoing and important process shaping eukaryotic genomes. The relative frequency of HGT events is very difficult to estimate. Most likely, the HGT events that researchers are discovering in eukaryotes are just "the tip of the iceberg". Indeed, the incorporation of foreign DNA into eukaryotic genomes is a complex process. If transfer events go to completion, only those sequences that provide some advantage to the new genome have a chance at preservation. Otherwise, they are quickly inactivated and eliminated from the nucleus, as is for example the case of chloroplast fragments that were incorporated into the nucleus of tobacco (Stegemann and Bock, 2009). This means that we are most likely to detect those events that are functional or that have been transferred recently, or those that have still remained some evolutionary signal detectable by current phylogenetic methods. Furthermore, the case of HGT involving mitochondrial genes reminds us that confirming that an HGT event has happened is difficult; one has to carefully consider other alternative explanations (Martin, 2005; Cusimano et al., 2008; Goremykin et al., 2009). Moreover, sampling is still very sparse; despite enormous advances in sequencing methods and comparative genomics, we are still at the very beginning.

Species-specific factors accounting for HGT— It is immensely interesting to follow the pattern of HGT occurrence that starts to emerge from the accumulating data on HGT. Despite the sparse sampling, it is possible to assume that species vary considerably in the propensity to undergo HGT. In contrast to the numerous events of natural HGT in tobacco and other solanaceous species, there are no reports of natural HGT in other plant species in literature. For example, despite the availability of its complete genome sequence, no HGT events have been detected in A*rabidopsis thaliana* (*Brassicaceae*). However, though considerable effort has been leveraged at detecting HGT events in eukaryotes, it remains possible that the

genomes of other completely sequenced plant genomes have not been sufficiently surveyed for foreign genes that could have originated via HGT. Detailed analysis is sometimes able to reveal cases of HGT that can otherwise escape the detection (e.g., Rice and Palmer, 2006). It is possible to speculate that detailed analysis of all the so far sequenced genomes could reveal new cases of HGT in land plants. Although such analyses would require substantial effort, we believe this would be a valuable initiative. We performed a series of BLAST comparisons of several viral genomes (all completely sequenced NCBI representatives of the family Bromovirideae, Caulimovirideae and Geminiviridae) and bacterial species able to infect *A. thaliana* (completely sequenced genomes of *Agrobacterium tumefaciens, Pseudomonas syringae, Pseudomonas aeruginosa, Ralstonia solanacearum, Xanthomonas campestris*, along with plasmid sequences of *Agrobacterium rhizogenes, A. radiobacter, A. vitis*) with the genome of *A. thaliana* (more details in Appendix S2, see Supplemental Data with the online version of this article). However, we did not detect any convincing traces of potential HGT involving these species. Nevertheless, our approach was limited to only three groups of viruses that are known to infect plants and few bacteria that have been shown to infect *A. thaliana* and most probably our methods would not have been able to detect more ancient HGT events. Undetected sequences notwithstanding, the seeming lack of the foreign sequences in *A. thaliana* genome is intriguing, and it is tempting to hypothesize about the reasons why it is so. If we relax the assumption that HGT occurrence is a purely random process, it would be worth examining the hypothesis that there are species-specific factors that could increase the probability of HGT events. We admit that in this regard comparing *Arabidopsis* with tobacco is like comparing David with Goliath. *Arabidopsis* is a rather ephemeral species with a rapid life cycle, and thus it is less likely to interact extensively with other species. On the other hand, tobacco lives longer (even though the life-span can vary within *N. tabacum*), giving it more potential opportunities for HGT. However, lifespan may

not be the best determinant of species interactions. One might assume, due to the ephemeral life strategy of *A. thaliana*, that it is less prone to pathogens, and that interactions with pathogens are less important in this species. However, this is not the case; many pathogens are able to seriously attack *A. thaliana*, and numerous resistance genes were selected in natural populations to fight bacterial and fungal pathogens (reviewed by Polland et al. 2009).

It is also well known that species like tobacco and its relatives can readily regenerate (in vitro even from a single somatic cells); moreover, they easily propagate clonally. Another noteworthy fact is that plastid transformation is to date commonly performed only in tobacco because the efficiency in other plants is very low (reviewed in Wang et al. 2009). We can speculate that the high degree of the genome plasticity combined with some specificities of the lifestyle of the species (e.g., diverse associations with other species, life-span, vulnerability to pathogens, etc.) could account for the likelihood that horizontal gene transfer will occur. In contrast, *A. thaliana* is known for its compact genome, suggesting that selection forces act at the genome level to maintain reduced genome size. In plants, DSB repair is thought to contribute to the variation of genome sizes; it prevents genome enlargement caused by the spread of retrotransposons (Grover et al., 2008). Since insertion of exogenous DNA has been though to occur via double-strand break repair (e.g., Blanchard and Schmidt, 1995; Salomon and Puchta, 1998;), species-specific differences in the effectiveness of DNA repair machinery could also contribute to the differences in the content of foreign DNA. Comparisons of deletion formation via double-strand break (DSB) repair in somatic cells between tobacco and *Arabidopsis thaliana* revealed strong differences between these two relatively closely related taxa with big differences in genome size (~20 fold) (Kirik et al., 2000). While in *Arabidopsis* DSB repair is predominantly accompanied by deletions, most probably resulting in its small genome size, in tobacco, various insertions into the break site along with smaller-scale deletions (in comparison with *Arabidopsis*) have been observed

(Kirik et al., 2000). In accordance with this, Orel and Puchta (2003) further showed that exonucleolytic degradation of exogeneous DNA is much more effective in *Arabidopsis* than in tobacco. Unfortunately, there are no data available for comparisons of other species within the same family. Therefore, the hypothesis that DNA repair mechanisms influence the likelihood of incorporation of foreign DNA into the genome is still questionable. It is most probably a complex of many other factors that is ultimately responsible for the difference in the content of foreign DNA.

Adaptive role and regulation of natural transgenes— Regarding the transfer of adaptive traits, bacteria have much more to offer compared to viruses. Nonetheless, gains in selective advantage from transferred genes are not well documented. *Rol* genes acquired from *A. rhizogenes* have been suggested to contribute to the phenotype of *Nicotiana* species, leading to reproductive isolation and speciation in the history of this genus (e.g., Suzuki et al., 2002). If this had been the case, we would expect the phenotypes characteristic of *rol* gene expression of would be exhibited by wild-type species. Unfortunately, to our knowledge, such traits have not yet been reported. The adaptive function of integrated sequences of viral origin and their impact on plant genomes is also currently unresolved. In most cases, endogenous viral sequences are present in numerous copies and have a repetitive character. Plants grown under normal growth conditions do not usually exhibit symptoms of infection. Detection of short RNAs homologous to endogenous pararetroviruses in healthy tomato plants (*Lyc*EPRV) and the methylation of *Lyc*EPRV (Staginnus et al., 2007) have confirmed previous hypotheses suggesting that EPRVs might confer resistance to the exogeneous virus (Richert-Pöggeler and Shepherd, 1997; Jakowitsch et al., 1999; Gregor et al., 2004). Thus the suggested adaptive role of these sequences could be connected with suppression of viral

infection by RNA-mediated silencing, i.e., RNA-directed DNA methylation, or postranscriptional- and translational gene silencing (PTGS and TGS) (Staginnus et al., 2007). Whatever the adaptive role, if any, of these insertions might be, their activity appears to be under strict control. The role of methylation-mediated regulation of expression of foreign inserts in plants is supported by the finding that their activity can be restored or increased during stress conditions. Such activation of both bacterial and viral insertions was observed mostly in interspecific hybrids and was intensified by abiotic stress (growing in in vitro cell cultures, frequent wounding, water stress, grafting, heat stress or changing light regime) (Aoki et al., 1994; Ahuja, 1996; Lockhart and Lesmann, 1997; Lockhart et al., 2000; Dallot et al., 2001; Zeidan et al., 2001; Richert-Pöggeler et al., 2003; Lhereux et al., 2003; Udagawa et al., 2004). The role of stress in HGT does not seem to be restricted to the expression regulation of inserts already present in the genome. Incorporation of foreign DNA might be promoted if host cells to lose their ability to guard their nuclear genomes (e.g., induction of double-strand breaks by transposon activation during stress conditions (e.g., Capy et al., 2000; Puchta, 2005)). If this is the case, it opens the question of whether stress-mediated transfers are controlled or random, and whether they somehow enable plants to deal with the stress. However, we currently lack the experimental data that could shed light on these issues.

***Agrobacterium is (not) alone*—** The notion that bacteria are the most common donors of genetic material to eukaryotes has been challenged by numerous findings of viral sequences in flowering plant genomes. Only a few cases of HGT from bacteria to flowering plants have been detected so far. They include *Agrobacterium rhizogenes* insertions (*rol* genes) found in *Nicotiana* species (Table 1). Can we assume that there might be other bacteria able to naturally transform their eukaryotic hosts? It is not unlikely, given that the determinants of gene transfer could be shared by horizontal gene transfer between bacteria. Broothaerts et al.

(2005) suggest that if there are other genes responsible for gene transfer other than those located on the Ti plasmid of *A. tumefaciens* (gene transfer being only a part of otherwise more complex virulence phenotype), other bacteria (at least *Rhizobia*) possess homologs of these genes. Moreover, Broothaerts et al. (2005) hypothesized that those few *vir*-genes present on the Ti plasmid could be sufficient to make bacteria competent for the gene transfer.

Few other bacteria (*Rhizobium, Sinorhizobium, Mesorhizobium*) related to *Agrobacterium* sp. were shown to be able to naturally transform plants as long as they have been equipped with the Ti plasmid of *Agrobacterium tumefaciens* (Broothaerts et al., 2005). There is also a report of a horizontal gene transfer from *Ralstonia solanacearum*-like bacterium into the potato (*Solanum tuberosum*) genome accompanied by protein domain reshuffling, which has resulted in a mosaic protein that confers resistance to *Ralstonia solanacearum* (Feng et al., 2003; Talianova et al., submitted).

DNA trafficking through cell-to-cell channels— Apart from a vector- mediated (e.g., viruses, bacteria or insects) DNA transfers, an alternative mechanisms of HGT relies on the ability of relatively small informational macromolecules (RNAs and proteins), and possibly also of the DNA fragments, to migrate between the cells as suggested by the tissue grafting experiments on transgenic tobacco. Stegemann and Bock (2009) hypothesized that the transfer of chloroplast DNA (or even the whole organelle) could occur through plasmodesmal connections of cells in the tobacco graft and scion junction. However, experiments conducted in the holoparasitic angiosperm genus *Pilostyles* have shown that whole, intact mitochondria are too large to be passed through plasmodesmata (Nickrent et al., 2004). More support for the hypothesis of Stegemann and Bock was provided by studies on associations of plant parasites such as shoot holoparasites *Cuscuta* sp. (*Convolvulaceae*) or root parasites *Phelipanche* sp. (*Orobanchaceae*) with their plant hosts. These studies have shown that there

is an intensive flow of pathogens, such as viruses (Hosford, 1967; Roos and Aldrich, 1988; Gal-On et al., 2009) or phytoplasmas (e.g., Heintz, 1989; Macrone et al., 1999), and macromolecules (proteins and RNAs) (Haupt et al., 2001; Roney et al., 2007; Westwood et al. 2009) between the host and parasite. Such findings indicate that the symplastic link (i.e., plasmodesmata) is established at the host-parasite interface to allow for such transport. Trafficking of macromolecules through plasmodesmata is selective and regulated, involving various plasmodesmal constituents, accompanied by an actino-myosin-dependent mechanism inside plasmodesmata (Gerdes et al., 2007; Lucas et al., 2009). Physical limitation (i.e., size exclusion limit; SEL) has been documented to be variable and several proteins have been identified that can induce increase the SEL and widen of plasmodesmal channels (Lucas et al., 2009). The plasmodesmal SEL is supposed to vary according to the developmental stage, and it is probably tissue- or even cell-type specific (Zambryski and Crawford, 2000; Cilia and Jackson, 2004; Ueki and Citovsky, 2005).

Whether bulk DNA or organelles can be transported in this manner remains undetermined. Interestingly, in mammals, *de novo* formation of membrane channels (i.e., tunneling nanotubes; TNT) similar to plasmodesmal channels in plants have been reported to occur in various cell-types (Gerdes and Carvalho, 2008). These channels have been shown to allow direct intercellular transfer of organelles (mitochondria or endosomes) (e.g., Koyanagi et al., 2005), components of plasmatic membrane (e.g., Rustom et al., 2004), or viral proteins (e.g., Sowinski et al., 2008).

In flowering plants, there are reports of the formation of cytoplasmic channels during a process called cytomixis - an infrequent cytological phenomenon that occurs in wide range of plant species (e.g., Heslop-Harisson, 1966). It involves the migration of chromatin material/ chromosomes through intercellular bridges. Cytomictic channels are thought to be derived from plasmodesmata (Heslop-Harrison, 1966) since they are large enough to enable transfer

of whole nuclei or cytoplasmic organelles (Risueno et al., 1969). Despite being observed at very low frequency in mitotic cells (e.g., Bowes, 1973; Wang, 2004), cytomixis most frequently occurs spontaneously during the meiosis of pollen mother cells. This often results in gametes with reduced fertility and increased or decreased ploidy levels and might contribute to evolution of aneuploid and polyploid plant species (e.g., Lattoo et al., 2006; Negrón-Ortiz, 2007; Singhal and Kumar, 2008; Mursalimov et al., 2010). However, due to its rarity and harmful effects, cytomixis should be regarded as a pathological process.

Based on above findings it appears that supracellularity applies for fungi, plants and animals (Baluška, 2009). Eukaryotic cells appear to communicate intensively and selectively among each other. Most interestingly, such connections can be established among cells of different organisms (e.g., host and parasite in plants), or they can even be forced by pathogens that enable them to spread. Whether this route has more often contributed to horizontal gene transfers in eukaryotes, however, remains to be clarified. In addition, in the studies of HGT, special attention should be given to the genes whose transcripts are present in the phloem sap, since they are accessible for putative vectors (insects or parasitic plants).

Horizontal DNA transfer and GMO— As final reminder, we discuss horizontal gene transfer and GMOs (genetically modified organisms). HGT has raised many concerns regarding the use of GMOs, especially GMO plants. Many of these concerns are related to the release of transgenes to soil and their uptake by soil microorganisms. However, to date, such risks have been found to be negligible (e.g., Sweet, 2009). Moreover, the spread of the antibiotic resistance genes from crops among bacteria does not constitute any selective advantage, because various resistance genes and other genetic determinants are already naturally present in the soil (Demanèche et al. 2008). Also, the possible risk of the fungus mediated HGT does not seems very dangerous as it appears to be relatively rare, at least according to the present

status of knowledge (reviewed in Davis et al. 2005), and the putative cases of this type of HGT involved mainly mitochondrial genes (e.g., Bergthorsson et al., 2003). However, this risk should be taken into account because of the hypothetical possibility of uncontrolled spread of transgenes via mycorrhizal systems (see above). There is, however, a much higher risk of pollen-mediated gene flow between crop GMO plants and its crop or wild relatives. Many species from the family Solanaceae are important crop species that do undergo monitoring to prevent such transgene escape (e.g., Celis et al., 2004; Rotino et al., 2005; Warwick et al., 2009). Since one of the containment strategies to prevent pollen-mediated transgene flow relies on transformation of chloroplasts (e.g., Bock, 2007; Moon et al., 2009), recent findings showing that genes can traffic at high frequency from chloroplast to the nucleus in tobacco have important implications for biotechnology. Moreover, a low frequency of chloroplast paternal leakage, and transmission of chloroplast-genome transgenes via pollen has been shown in tobacco (Ruf et al., 2007). Even though solanaceous species appear to be very plastic, the risk of transgene flow mediated by HGT can be regarded as negligible compared to that mediated by pollen dispersal. Since there are numerous applications where the risk of transgene flow via pollen has to be reduced to zero, the above findings suggest that chloroplast transformation itself is not a panacea and should be accompanied by other containment measures (e.g., Warwick et al., 2009; Moon et al., 2009). To ascertain the level of such risks, more quantitative studies regarding the frequency of chloroplast DNA trafficking to the nucleus are needed from other species used for genetic modification.

In any case, discovering new routes of gene trafficking between species and understanding the mechanisms of HGT is one of the important prerequisites for effective transgene control in genetically modified organisms.

***Conclusion*—** The evolutionary history of plants compared to many animal and fungal taxa has been frequently accompanied by reticulate evolutionary events. Such reticulations are mainly caused by interspecific hybridization - a feature characteristic for plant kingdom. Horizontal gene transfer can also be regarded as a source of network-like evolution (not only in plants), however this process is much less frequent. With an advance in sequencing and comparative genomics, increasing amounts of data can be used not only for more exact reconstructions of tree-of-life, but also for identification of new cases of HGT. Due to the lack of experimental data, currently these events are mostly detected accidentally by studying their effects (e.g., incongruence in evolutionary reconstructions). Despite the scarcity of natural HGT cases and the lack of detailed knowledge about the mechanisms governing uptake and incorporation of foreign DNA in plants, some trends are already starting to come out. Studies on natural HGT in solanaceous species along with experimental studies on gene transfer using transgenic tobacco have provided researchers with interesting clues about novel routes to the spread of informational molecules, transgenes, organelles or pathogens, as well as clues about the incorporation and control of expression of natural transgenes in plant genomes.

Literature cited

ADAMS K. L., D. O. DALEY, Y. L. QIU, J. WHELAN, AND J. D. PALMER 2000. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* 408: 354-357.

AHUJA, M. R. 1962. A cytogenetic study of heritable tumors in Nicotiana species hybrids. *Genetics* 47: 865-880.

AHUJA, M. R. 1996. Genetic nature of a nontumour mutant isolated from tumour-prone amphidiploid Nicotiana glauca-langsdorffii (GGLL): a critical assessment. *Heredity* 76: 335-345.

ANDERSSON, J. O. 2005. Lateral gene transfer in eukaryotes. *Cellular and Molecular Life Sciences* 62: 1182-1197.

AOKI, S. 2004. Resurrection of an ancestral gene: functional and evolutionary analyses of the Ngrol genes transferred from Agrobacterium to Nicotiana. *Journal of Plant Research* 117: 329-337.

AOKI, S., A. KAWAOKA, M. SEKINE, T. ICHIKAWA, T. FUJITA, A. SHINMYO AND K. SYONO. 1994. Sequence of the cellular T-DNA in the untransformed genome of Nicotiana glauca that is homologous to ORFs 13 and 14 of the Ri plasmid and analysis of its expression in genetic tumors of N. glauca x N. langsdorffii. *Molecular and General Genetics* 243: 706-710.

AOKI, S., AND K. SYONO. 1999a. Function of Ngrol genes in the evolution of Nicotiana glauca: conservation of the function of NgORF13 and NgORF14 after ancient infection by an Agrobacterium rhizogenes-like ancestor. *Plant and Cell Physiology* 40: 222-230.

AOKI, S., AND K. SYONO. 1999b. Horizontal gene transfer and mutation: Ngrol genes in the genome of Nicotiana glauca. *Proceedings of the National Academy of Sciences, USA* 96: 13229-13234.

ARIMURA, S., J. YAMAMOTO, G. P. AIDA, M. NAKAZONO, AND N. TSUTSUMI. 2004. Frequent fusion and fission of plant mitochondria with unequal nucleoid distribution. *Proceedings of the National Academy of Sciences, USA* 101: 7805-7808.

ASHBY, M. K., A. WARRY, E. R. BEJARANO, A. KHASHOGGI, M. BURRELL, AND C.P. LICHTENSTEIN. 1997. Analysis of multiple copies of geminiviral DNA in the genome of four closely related Nicotiana species suggest a unique integration event. *Plant Molecular Biology* 35: 313-321.

BALUŠKA, F. 2009. Cell-cell channels, viruses, and evolution via infection, parasitism, and symbiosis toward higher levels of biological complexity. *Annals of the New York Academy of Sciences* 1178:106-119.

BARKMAN, T. J., J. R. MCNEAL, S.-H. LIM, G. COAT, H. B. CROOM, N. D. YOUNG, AND C. W. DEPAMPHILIS. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evolutionary Biology* 4: 248.

BEJARANO, E. R., A. KHASHOGGI, M. WITTY, AND C. LICHTENSTEIN. 1996. Integration of multiple repeats of geminivviral DNA into the nuclear genome of tobacco during evolution. *Proceedings of the National Academy of Sciences, USA* 93: 759-764.

BERGTHORSSON, U., K. L. ADAMS, B. THOMASON, AND J. D. PALMER. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424: 197-201.

BERGTHORSSON, U., A. O. RICHARDSON, G. J. YOUNG, L. R. GOERTZEN, AND J. D. PALMER. 2004. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm Amborella. *Proceedings of the National Academy of Sciences, USA* 101: 17747-17752.

BERTSCH, C., M. BEUVE, V. V. DOLJA, M. WIRTH, F. PELSY, E. HERRBACH, AND O. LEMAIRE. 2009. Retention of the virus-derived sequences in the nuclear genome of grapevine as a potential pathway to virus resistance. *Biology Direct* 4: 21

BLANCHARD, J. L., AND G. W. SCHMIDT. 1995. Pervasive migration of organellar DNA to the nucleus in plants. *Journal of Molecular Evolution* 41: 397-406.

BOCK, R. 2007. Plastid biotechnology: prospects for herbicide and insect resistance, metabolic engineering and molecular farming. *Current Opinion in Biotechnology* 18: 100-106.

BOCK, R. 2010. The give-and-take of DNA: horizontal gene transfer in plants. *Trends in Plant Science* 15: 11-22.

BOCK, R., AND J. N. TIMMIS. 2008. Reconstructing evolution: gene transfer from plastids to the nucleus. *BioEssays* 30: 556-566.

BOWES, B. G. 1973. Notes on apparent cases of cytomixis in the root apex of Allium cepa. *Cytologia* 38: 125-129.

BROOTHAERTS, W., H. J. MITCHELL, B. WEIR, S. KAINES, L. M. SMITH, W. YANG, J. E. MAYER, ET AL. 2005. Gene transfer to plants by diverse species of bacteria. *Nature* 433: 629-633.

CAPY, P., G. GASPERI, C. BIÉMONT, AND C. BAZIN. 2000. Stress and transposable elements: co-evolution or useful parasites? *Heredіy* 85: 101-106.

CELIS, C., M. SCURRAH, S. COWGILL, S. CHUMBIAUCA, J. GREEN, J. FRANCO, G. MAIN, ET AL. 2004. Environmental biosafety and transgenic potato in a cetnre of diversity for this crop. *Nature* 432: 222-225.

CHEN, S.-CH., H.-W. LIU, K.-T. LEE, AND T. YAMAKAWA. 2007. High-efficiency Agrobacterium rhizogenes mediated transformation of heat inducible sHSP18.2-GUS in Nicotiana tabacum. *Plant Cell Reports* 26: 29-37.

CHENG, X., D. ZHANG, Z. CHENG, B. KELLER, AND H.-Q. LING. 2009. A new family of Ty1-copia-like retrotransposons originated in the tomato genome by a recent horizontal transfer event. *Genetics* 181: 1183-1193.

CHO, Y., Y.-L. QIU, P. KUHLMAN, AND J. D. PALMER. 1998. Explosive invasion of plant mitochondria by a group I intron. *Proceedings of the National Academy of Sciences, USA* 95: 14244-14249.

CILIA, M.L., AND D. JACKSON. 2004. Plasmodesmata form and function. *Current Opinion in Cell Biology* 16: 500-506.

CUSIMANO, N., L.-B. ZHANG, AND S. S. RENNER. 2008. Reevaluation of the cox1 group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. *Molecular Biology and Evolution* 25: 265-276.

DALLOT, S., P. ACUNA, C. RIVERA, P. RAMIREZ, F. COTE, M. E. L. LOCKHART, AND M. L. CARUANA. 2001. Evidence that the proliferation stage of micropropagation procedure is determinant in the expression of Banana streak virus integrated into the genome of FHIA 21 hybrid (Musa AAAB). *Archives of Virology* 146: 2179-2190.

DARWIN, C. 1883. The variation of animals and plants under domestication. D. Appleton & Co., New York, USA.

DAVID, C., M. D. CHILTON, AND J. TEMPÉ. 1984. Conservation of T-DNA in plants regenerated from hairy root cultures. *Biotechnology* 2: 73-76.

DAVIS, C. C., AND K. J. WURDACK. 2004. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpigiales. *Science* 305: 676-678.

DAVIS, C.C., W. R. ANDERSON, AND K. J. WURDACK. 2005. Gene transfer from a parasitic flowering plant to a fern. *Proceedings of the Royal Society B: Biological Sciences* 272: 2237-2242.

DAY, A. G., E. R. BEJARANO, K. W. BUCK, M. BURRELL, AND C. P. LICHTENSTEIN. 1991. Expression of an antisense viral gene in transgenic tobacco confers resistance to the DNA virus tomato golden mosaic virus. *Proceedings of the National Academy of Sciences, USA* 88: 6721-6725.

DEMANÈCHE, S., H. SANGUIN, J. POTÉ, E. NAVARRO, D. BERNILLON, P. MAVINGUI, W. WILDI, ET AL. 2008. Antibiotic-resistant soil bacteria in transgenic plant fields. *Proceedings of the National Academy of Sciences, USA* 105: 3957-3962.

DIAO, X., M. FREELING, AND D. LISCH. (2006) Horizontal transfer of a plant transposon. *PLoS Biology* 4: e5

DURANTE, M., E. CECCHINI, L. NATALI, L. CITTI, C. GERI, R. PARENTI, AND V. N. RONCHI. 1989. 5-Azacytidine-induced tumorous transformation and DNA hypermethylation in Nicotiana tissue cultures. *Developmental Genetics* 10: 298-303.

FENG, J., F. YUAN, Y. GAO, C. LIANG, J. XU, C. ZHANG, AND L. HE. 2003. A novel antimicrobial protein isolated from potato (Solanum tuberosum) shares homology with an acid phosphatase. *Biochemical Journal* 376: 481-487.

FRUNDT, C., A. D. MEYER, T. ICHIKAWA, AND F. JR. MEINS (1998) A tobacco homologue of the Ri-plasmid orf13 gene causes cell proliferation in carrot root discs. *Molecular and General Genetics* 259: 559-568.

FURNER, I. J., G. A. HUFFMAN, R. M. AMASINO, D. J. GARFINKEL, M. P. GORDON, AND E. W. NESTER. 1986. An Agrobacterium transformation in the evolution of the genus Nicotiana. *Nature* 319: 422-427.

GAL-ON, A., A. NAGLIS, D. LEIBMAN, H. ZIADNA, K. KATHIRAVAN, L. PAPAYIANNIS, V. HOLDENGREBER, ET AL. 2009. Broomrape can acquire viruses from its hosts. *Virology* 99: 1321-1329.

GELVIN, S.B. 2010. Finding a way to the nucleus. *Current Opinion in Microbiology* 13: 53-58.

GERDES, H.H., N. V. BUKORESHTLIEV, AND J. F. V. BAROOSO. 2007. Tunneling nanotubes: A new route for the exchange of components between animal cells. *FEBS Letters* 581: 2194-2201.

GERDES, H. H., AND R. N. CARVALHO. 2008. Intercellular transfer mediated by tunneling nanotubes. *Current Opinion in Cell Biology* 20: 470-475.

GHATNEKAR, L., M. JAAROLA, AND B. O. BENGTSSON. 2006. The introgression of a functional nuclear gene from Poa to Festuca ovina. *Proceedings of the Royal Society B: Biological Sciences* 273: 395-399.

GLADYSHEV, E. A., M. MESELSON, AND I. R. ARKHIPOVA. 2008. Massive horizontal gene transfer in bdelloid rotifers. *Science* 320: 1210-1213.

GOODSPEED, T. H. 1954. The Genus Nicotiana. Chronica Botanica Company, Waltham, Massachusetts, USA.

GOREMYKIN, V.V., F. SALAMINI, R. VELASCO, AND R. VIOLA. 2009. Mitochondrial DNA of Vitis vinifera and the issue of rampant horizontal gene transfer. *Molecular Biology and Evolution* 26: 99-110.

GREGOR, W., M. F. METTE, C. STAGINNUS, M. A. MATZKE, AND A. J. M. MATZKE. 2004. A distinct endogenous pararetrovirus family in Nicotiana tomentosiformis, a diploid progenitor of polyploid tobacco. *Plant Physiology* 134: 1191-1199.

GROVER, C. E., J. S. HAWKINS, AND J. F. WENDEL. 2008. Phylogenetic insights into the pace and pattern of plant genome size evolution. *Genome Dynamics* 4: 57-68.

HANSEN, C. N., G. HARPER, AND J. S. HESLOP-HARRISON. 2005. Characterisation of pararetrovirus-like sequences in the genome of potato (Solanum tuberosum). *Cytogenetic and Genome Research* 110: 559-565.

HARPER, G., J. O. OSUJI, J. S. HESLOP-HARRISON, AND R. HULL. 1999. Integration of banana streak badnavirus into the Musa genome: molecular and cytogenetic evidence. *Virology* 255: 207-213.

HARPER, G., R. HULL, B. LOCKHART, AND N. OLSZEWSKI. 2002. Viral sequences integrated into plant genomes. *Annual Review of Phytopathology* 40: 119-136.

HAUPT, S, K. J. OPARKA, N. SAUER, AND S. NEUMANN. 2001. Macromolecular trafficking between Nicotiana tabacum and the holoparasite Cuscuta reflexa. *Journal of Experimental Botany* 52: 173-177.

HEINTZ, W. 1989. Transmission of a new mycoplasma-like organism (MLO) from Cuscuta odorata (Ruiz et Pav.) to herbaceous plants and attempts to its elimination in the vector. *Journal of Phytopathology* 125: 171-186.

HESLOP-HARRISON, J. 1966. Cytoplasmic connexions between angiosperm meiocytes. *Annals of Botany* 30: 592-600

HOSFORD, R. M. 1967. Transmission of plant viruses by dodder. *The Botanical Review* 33: 387-406.

HOTOPP, J. C. D., M. E. CLARK, D. C. S. G. OLIVEIRA, J. M. FOSTER, P. FISCHER, M. C. M. TORRES, J. D. GIEBEL, ET AL. 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317: 1753-1756.

HUANG, C. Y., M. A. AYLIFFE, AND J. N. TIMMIS. 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* 422: 72-76.

HUANG, C. Y., M. A. AYLIFFE, AND J. N. TIMMIS. 2004. Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco. *Proceedings of the National Academy of Sciences, USA* 101: 9710-9715.

INTRIERI, M. C., AND M. BUIATTI. 2001. The horizontal transfer of Agrobacterium rhizogenes genes and the evolution of the genus Nicotiana. *Molecular Phylogenetics and Evolution* 20: 100-110.

JAKOWITSCH, J., M. F. METTE, J. VAN DER WINDEN, M. A. MATZKE, AND A. J. M. MATZKE. 1999. Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. *Proceedings of the National Academy of Sciences, USA* 96: 13241-13246.

JELÍNKOVÁ, H., F. TREMBLAY, AND A. DESROCHERS. 2009. Molecular and dendrochronological analysis of natural root grafting in *Populus tremuloides* (Salicaceae). *American Journal of Botany* 96: 1500-1505.

JONES, J. J., C. FURLANETTO, AND T. KIKUCHI. 2005. Horizontal gene transfer from bacteria and fungi as a driving force in the evolution of plant parasitism in nematodes. *Nematology* 7: 641-646.

KARAMI, O., M. ESNA-ASHARI, G. K. KURDISTANI, AND B. AGHAVAISI. 2009. Agrobacterium-mediated genetic transformation o plants: the role of host. *Biologia Plantarum* 53: 201-212.

KEJNOVSKY, E., I. J. LEITCH, AND A. R. LEITCH. 2009. Contrasting evolutionary dynamics between angiosperm and mammalian genomes. *Trends in Ecology and Evolution* 24: 572-582.

KEELING, P. J. 2009. Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Current Opinion in Genetics and Development* 19: 613-619.

KEELING, P. J. 2010. The endosymbiotic origin, diversification and fate of plastids. *Philosophical Transactions of the Royal Society B* 365: 729 -748.

KEELING, P. J., AND J. D. PALMER. 2008. Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics* 9: 605-618

KOULINTCHENKO, M., Y. KONSTANTINOV, AND A. DIETRICH. 2003. Plant mitochondria actively import DNA via the permeability transition pore complex. *The EMBO Journal* 22: 1245-1254.

KIRIK, A., S. SALOMON, AND H. PUCHTA. 2000. Species-specific double-strand break repair and genome evolution in plants. *The EMBO Journal* 19: 5562-5566.

KOYANAGI, M., R. P. BRANDES, J. HAENDELER, A. M. ZEIHER, AND S. DIMMELER. 2005. Cell-to-cell connection of endothelial progenitor cells with cardiac myocytes by nanotubes: a novel mechanism for cell fate changes? *Circulation Research* 96: 1039-1041.

KUNII, M., M. KANDA, H. NAGANO, I. UYEDA, Y. KISHIMA, AND Y. SANO. 2004. Reconstruction of putative DNA virus from endogenous rice tungro bacilliform virus-like sequences in the rice genome: implications for integration and evolution. *BMC Genomics* 5: 80.

LATTOO, S. K., S. KHAN, S. BAMOTRA, A. K. DHAR. 2006. Cytomixis imparis meiosis and influences reproductive success in Chlorophytum comosum (Thunb) Jacq. – an additional strategy and possible implications. *Journal of Biosciences* 31: 629-637.

LAZAROWITZ, S. G. 1992. Geminiviruses: genome structure and gene function. *Critical Reviews in Plant Sciences* 11: 327-349.

LEE, J.-M. AND M. ODA 2003. Grafting of herbaceuos vegetable and ornamental crops. *In* J. Janick [ed.] Horticultural Reviews 28, 61-125, John Wiley & Sons, New York, New York USA.

LEISTER, D. 2005. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends in Genetics* 21:655-663.

LHEUREUX, F., F. CARREEL, C. JENNY, B. E. L. LOCKHART, AND M. L. ISKRA-CARUANA. 2003. Identification of genetic markers linked to Banana streak disease expression in inter-specific Musa hybrids. *Theoretical and Applied Genetics* 106: 594-598.

LIU, Y. 2006. Historical and modern genetics of plant graft hybridization. *Advances in Genetics* 56: 101-129.

LOCKHART, B. E. L., AND D. E. LESEMANN. 1997. Occurrence of petunia vein clearing virus in the U.S.A. *Plant Disease* 82: 262.

LOCKHART, B. E., J. MENKE, G. DAHAL, AND N. E. OLSZEWSKI. 2000. Characterization and genomic analysis of tobacco vein clearing virus, a plant pararetrovirus that is transmitted vertically and related to sequences integrated in the host genome. *Journal of General Virology* 81: 1579 – 1585.

LUCAS, W. J., B.-K. HAM, AND J.-Y. KIM. 2009. Plasmodesmata – bridging the gap between neighboring plant cells. *Trends in Cell Biology* 19: 495-503.

MACRONE, C., F. HERGENHAHN, A. RAGOZZINO, AND E. SEEMÜLLER. 1999. Dodder transmission of pear decline, european stone fruit yellows, rubus stunt, picris echoides yellows and cotton phyllody phytoplasmas to periwrinkle. *Journal of Phytopathology* 147: 187-192.

MARTIN, W. 2005. Lateral gene transfer and other possibilities. *Heredity* 94: 565-566.

MATZKE, M., W. GREGOR, M. F. METTE, W. AUFSATZ, T. KANNO, J. JAKOWITSCH, AND A. J. M. MATZKE. 2004. Endogenous pararetroviruses of allotetraploid Nicotiana tabacum and its diploid progenitors, N. sylvestris and N. tomentosiformis. *Biological Journal of the Linnean Society* 82: 627-638.

RUIZ-MEDRANO, R., XOCONOSTLE-CÁZARES, B, LUCAS, W.J. 1999. Phloem long-distance transport of CmNACP mRNA: implications for supracellular regulation in plants. *Development* 126:4405-4419.

METTE, M. F., T. KANNO, W. AUFSATZ, J. JAKOWITSCH, J. VAN DER WINDEN, M. A. MATZKE, AND A. J. M. MATZKE. 2002. Endogenous viral sequences and their potential contribution to heritable virus resistance in plants. *The EMBO Journal* 21: 461-469.

MEYER, A. D., T. ICHIKAWA, AND F. JR. MEINS. 1995. Horizontal gene transfer: regulated expression of a tobacco homologue of the Agrobacterium rhizogenes rolC gene. *Molecular and General Genetics* 249: 265-273.

MOON, H. S., Y. LI, AND C. N. JR. STEWART. 2009. Keeping the genie in the bottle: transgene biocontainment by excision in pollen. *Trends in Biotechnology* 28: 3-8.

MOWER, J. P., S. STEFANOVIC, G. J. YOUNG, AND J. D. PALMER. 2004. Gene transfer from parasitic to host plants. *Nature* 432: 165-166.

MURAD, L., K. Y. KIM, V. CHRISTOPODULOU, R. MATYASEK, C. P. LICHTENSTEIN, AND A. KOVARIK. 2002. The origin of tobacco's T genome is traced to a particular lineage within Nicotiana tometosiformis (Solanaceae). *American Journal of Botany* 89: 921-928.

MURAD, L., J. P. BIELAWSKI, R. MATYASEK, A. KOVARIK, R. A. NICHOLS, A. R. LEITCH, AND C. P. LICHTENSTEIN. 2004. The origin and evolution of geminivirus-related DNA sequences in Nicotiana. *Heredity* 92: 352-358.

MURSALIMOV, S. R., S. I. BAIBORODIN, Y. V. SIDORCHUK, V. K. SHUMMY, AND E. V. DEINEKO. 2010. Characteristics of the cytomictic channel formation in Nicotiana tabacum L. pollen mother cells. *Cytology and Genetics* 44: 14-18.

NAGATA, N., S. KOSONO, M. SEKINE, A. SHINMYO, AND K. SYONO. 1995. The regulatory functions of the rolB and rolC genes of Agrobacterium rhizogenes are conserved in the homologous genes (Ngrol) of Nicotiana glauca in tobacco genetic tumors. *Plant and Cell Physiology* 36: 1003-1012.

NAGATA, N., C. SAITO, A. SAKAI, H. KUROIWA, AND T. KUROIWA. 1999. The selective increase or decrease of organellar DNA in generative cells just after pollen mitosis one controls cytoplasmic inheritance. *Planta* 209: 53-65.

NEGRÓN-ORTIZ, V. 2007. Chromosome numbers, nuclear DNA content, and polyploidy in Consolea (Cactaceae), an endemic cactus of the Caribbean islands. *American Journal of Botany 94:* 1360-1370.

NDWORA, T., G. DAHAL, D. LAFLEUR, G. HARPER, R. HULL, N. E. OLSZEWSKI, AND B. LOCKHART. 1999. Evidence that badnavirus infection in Musa can originate from integrated pararetroviral sequences. *Virology* 255: 214 – 220.

NICKRENT, D. L., A. BLARER, Y. L. QIU, R. VIDAL-RUSSELL, AND F. E. ANDERSON. 2004. Phylogenetic inference in Rafflesiales: the influence of rate heterogeneity and horizontal gene transfer. *BMC Evolutionary Biology* 4: 40.

NOUTSOS, C., E. RICHLY, AND D. LEISTER. 2005. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. *Genome Research* 15: 616-628.

NUGENT, J. M., AND J. D. PALMER. 1991. RNA-mediated transfer of the gene coxII from the mitochondrion to the nucleus during flowering plant evolution. *Cell* 66: 473-481.

OREL, N., AND H. PUCHTA. 2003. Differences in the processing of DNA ends in *Arabidopsis thaliana* and tobacco: possible implications for genome evolution. *Plant Molecular Biology* 51: 523-531.

PAHALAWATTA, V., K. DRUFFEL, AND H. PAPPU. 2008. A new and distinct species in the genus Caulimovirus exist as an endogenous plant pararetroviral sequence in its host, Dahlia variabilis. *Virology* 376: 253-257.

POLAND, J.A., BALINT-KURTI, P.J., WISSER, R.J., PRATT, R.C., and R.J. NELSON.2009. Shades of gray: the world of quantitative disease resistance. *Trends in Plant Science* 14: 21-29.

PUCHTA, H. 2005. The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *Journal of Experimental Botany* 56: 1-14.

REDEI, G. P. 1998. Genetics Manual: Current Theory, Concepts, Terms. World Scientific, Singapore.

RICHARDS, T. A., D. M. SOANES, P. G. FOSTER, G. LEONARD, C. R. THORNTON, AND N. J. TALBOT. 2009. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *The Plant Cell* 21: 1897-1911.

RICHARDS, E.J. 2006. Inherited epigenetic variation — revisiting soft inheritance. *Nature Reviews Genetics* 7: 395-401.

RICHARDSON, A. O., AND J. D. PALMER. 2007. Horizontal gene transfer in plants. *Journal of Experimental Botany* 58: 1-9.

RICHERT-PÖGGELER, K. R., AND R. J. SHEPHERD. 1997. Petunia vein-clearing virus: a plant pararetrovirus with the core sequences for and integrase function. *Virology* 236: 137-146.

RICHERT-PÖGGELER, K. R., F. NOREEN, T. SCHWARZACHER, G. HARPER, AND T. HOHN. 2003. Induction of infectious petunia vein clearing (pararetro) virus from endogenous provirus in petunia. *EMBO Journal* 22: 4836-4845.

RICE, D.W., AND J. D. PLAMER. 2006. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biology* 4: 31.

RISUENO, M. C., G. GIMENEZ-MARTIN, J. F. LOPEZ-SAEZ, AND M. I. R. GARCIA. 1969. Connexions between meiocytes in plants. *Cytologia* 34: 262-272.

DE LA RIVA, G. A., J. GONZALES-CABRERA, R. VAZQUEZ-PADRON, AND C. AYRA-PARDO. 1998. Agrobacterium tumefaciens: a natural tool for plant transformation. *Electronic Journal of Biotechnology* 1: 118-133.

RONEY, J. K., P. A. KHATIBI, AND J. H. WESTWOOD. 2007. Cross-species translocation of mRNA from host plants into the parasitic plant dodder. *Plant Physiology* 143: 1037-1043.

ROOS, U. P., AND H. C. ALDRICH. 1988. Intracellular localization and morphology of rhabdovirus-like particles in dodder (Cuscuta sp.). *Canadian Journal of Botany* 66: 2499-2504.

ROSEWICH, U. L., AND H. C. KISTLER. 2000. Role of horizontal gene transfer in the evolution of fungi. *Annual Review of Phytopathology* 38: 325-363.

ROTINO, G. L., N. ACCIARRI, E. SABATINI, G. MENNELLA, R. L. SCALZO, A. MAESTRELLI, B. MOLESINI, ET AL. 2005. Open field trail of genetically modified parthenocarpic tomato: seedlessness and fruit quality. *BMC Biotechnology* 5: 32.

ROULIN, A., B .PIEGU, R. A. WING, AND O. PANAUD. 2008. Evidence of multiple horizontal transfers of the long terminal repeat retrotransposon RIRE1 within the genus Oryza. *The Plant Journal* 53: 950-959.

ROULIN, A., B. PIEGU, P. M. FORTUNE, F. SABOT, A. D'HONT, D. MANICACCI, AND O. PANAUD. 2009. Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfer of the LTR-retrotransposon Route66 in Poaceae. *BMC Evolutionary Biology* 9: 58.

RUSTOM, A., R. SAFFRICH, I. MARKOVIC, P. WALTHER, AND H. H. GERDES. 2004. Nanotubular highways for intercellular organelle transport. *Science* 303: 1007-1010.

RUF S., D. KARCHER, AND R. BOCK. 2007. Determining the transgene containment level provided by chloroplast transformation. *Proceedings of the National Academy of Sciences, USA* 104: 6998-7002.

SALOMON, S., AND H. PUCHTA. 1998. Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO Journal* 17: 6086-6095.

SANCHEZ-PUERTA, M.V., Y. CHO, J. P. MOWER, A. J. ALVERSON, AND J. D. PALMER. 2008. Frequent, phylogenetically local horizontal transfer of the cox1 group I intron in flowering plant mitochondria. *Molecular Biology and Evolution* 25: 1762-1777.

SCHÖNENBERGER, J., A. A. ANDERBERG, K. J. SYTSMA. 2005. Molecular phylogenetics and patterns of floral evolution in the Ericales. *International Journal of Plant Sciences* 166: 265-288.

SHEPPARD, A. E., M. A. AYLIFFE, L. BLATCH, A. DAY, S. K. DELANEY, N. KHAIRUL-FAHMY, Y. LI, ET AL. 2008. Transfer of plastid DNA to the nucleus is elevated during male gametogenesis in tobacco. *Plant Physiology* 148: 328-336.

SINGHAL, V. K., AND P. KUMAR. 2008. Impact of cytomixis on meiosis, pollen viability and pollen size in wild populations of Himalayan poppy (Meconopsis aculeata Royle). *Journal of Biosciences* 33: 371-380.

SOWINSKI, S., C. JOLLY, O. BERNINGHAUSEN, M. A. PURBHOO, A. CHAUVEAE, K. KOHLER, S. ODDOS, ET AL. 2008. Membrane nanotubes physically connect T cells over long distances presenting a novel route for HIV-1 transmission. *Nature Cell Biology* 10: 211-219.

SMITH, H. H. 1988. The inheritance of genetic tumors in Nicotiana hybrids. *The Journal of Heredity* 79: 277-283.

SMITH S.E., AND D.J. READ. 1997. Mycorrhizal Symbiosis, 2nd ed. Academic Press, London, UK.

STAGINNUS C., AND K. R. RICHERT-PÖGGELER. 2006. Endogenous pararetroviruses: two-faced travelers in the plant genome. *TRENDS in Plant Science* 11: 485-491.

STAGINNUS, C., W. GREGOR, M. F. METTE, C. H. TEO, E. G. BORROTO-FERNANDEZ, M. L. C. MACHADO, M. MATZKE, AND T. SCHWARZACHER. 2007. Endogenous pararetroviral sequences in tomato (Solanum lycopersicum) and related species. *BMC Plant Biology* 7: 24.

STEGEMANN, S., S. HARTMANN, S. RUF, AND R. BOCK. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proceedings of the National Academy of Sciences, USA* 100: 8828-8833.

STEGEMANN, S., AND R. BOCK. 2006. Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *The Plant Cell* 18: 2869-2878.

STEGEMANN, S., AND R. BOCK. 2009. Exchange of genetic material between cells in plant tissue grafts. *Science* 324: 649-651.

SU L., S. GAO, Y. HUANG, C. JI, D. WANG, Y. MA, R. FANG, AND X. CHEN. 2007. Complete genomic sequence of Dracaena mottle virus, a distinct badnavirus. *Virus Genes* 35: 423-429.

SUZUKI, K., I. YAMASHITA, AND N. TANAKA. 2002. Tobacco plants were transformed by Agrobacterium rhizogenes infection during their evolution. *The Plant Journal* 32: 775-787.

SWEET, J. 2009. The 10th International Symposium on the Biosafety of Genetically Modified Organisms (ISBGMO), Wellington, New Zealand, November 2008. *Environmental Biosafety Research* 8: 161-181.

TALLER, J., Y. HIRATA, N. YAGISHITA, M. KITA, AND S. OGATA. 1998. Graft-induced genetic changes and the inheritance of several characteristics in pepper (Capsicum annuum L.) *Theoretical and Applied Genetics* 97: 705-713.

TANNE, E., AND I. SELA. 2005. Occurrence of a DNA sequence of a non-retro RNA virus in a host plant genome and its expression: evidence for recombination between viral and host RNAs. *Virology* 332: 614-622.

TEPFER, D. 1984. Transformation of several species of higher plants by Agrobacterium rhizogenes: sexual transmission of the transferred genotype and phenotype. *Cell* 37: 959-967.

TIMMIS, J. N., M. A. AYLIFFE, C. Y. HUANG, AND W. MARTIN. 2004. Endosymbotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* 5: 123-135.

TZFIRA, T., AND V. CITOVSKY. 2006. Agrobacterium-mediated genetic transformation of plants: biology and biotechnology. *Current Opinion in Biotechnology* 14:147-154.

UDAGAWA, M., S. AOKI, AND K. SYONO. 2004. Expression analysis of the NgORF13 promoter during the development of tobacco genetic tumors. *Plant and Cell Physiology* 45: 1023-1031.

UEKI, S., AND V. CITOVSKY. 2005. Control improves with age: Intercellular transport in plant embryos and adults. *Proceedings of the National Academy of Sciences, USA* 102: 1817-1818.

VAUGHN, J. C., M. T. MASON, G. L. SPER-WHITIS, P. KUHLMAN, AND J. D. PALMER. 1995. Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric CoxI gene of Peperomia. *Journal of Molecular Evolution* 41: 563-572.

VISSER, R. G. F., E. JACOBSEN, A. HESSELING-MEINDERS, M. J. SCHANS, B. WITHOLT, AND W. J. FEENSTRA. 1989. Transformation of homozygous diploid potato with an Agrobacterium tumefaciens binary vector system by adventitious shoot regeneration on leaf and stem segments. *Plant Molecular Biology* 12: 329-337.

WANG, B., AND Y.L. QIU 2006. Phylogenetic distribution and evolution of mycorrhizas in land plants. Mycorrhiza 16: 299–363.

WANG, S. Y., C. K. YUI, S. LEE, C. I. WANG, AND G. C. CHZAN. 2004. Ultrastructure and possible origin of the cytoplasmic channels providing connection between cells of anther vegetative tissues. *Russian Journal of Plant Physiology* 51: 110-120.

WANG, H. H., W. B. YIN, AND Z. M. HU. 2009. Advances in chloroplast engineering. *Journal of Genetics and Genomics* 36: 387-398.

WARWICK, S. I., H. J. BECKIE, AND L. M. HALL. 2009. Gene flow, invasiveness, and ecological impact of genetically modified crops. The Year in Evolutionary Biology 2009: *Annals of the New York Academy of Sciences* 1168: 72 – 99.

WESTWOOD, J.H., RONEY, J.K., KHATIBI, P.A., AND V.K STROMBERG. 2009. RNA translocation between parasitic plants and their hosts. *Pest Management Science* 65, 533-539.

WHITE, F. F., D. J. GARFINKEL, G. A. HUFFMAN, M. P. GORDON, AND E. W. NESTER. 1985. Molecular and genetic analysis of the transferred DNA regions of the root-inducing plasmid of Agrobacterium rhizogenes. *Journal of Bacteriology* 164: 33-44.

WOLOSZYNSKA, M., T. BOCER, P. MACKIEWICZ, AND H. JANSKA. 2004. A fragment of chloroplast DNA was transferred horizontally, probably from non-eudictos, to mitochondrial genome of Phaseolus. *Plant Molecular Biology* 56: 811-820.

WON, H., AND S. S. RENNER. 2003. Horizontal gene transfer from flowering plants to Gnetum. *Proceedings of the National Academy of Sciences, USA* 100: 10824-10829.

YANG, Z.-N., X.-R. YE, J. MOLINA, M. L. ROOSE, AND T. E. MIRKOV. 2003. Sequence analysis of a 282-kilobase region surrounding the Citrus tristeza resistance gene (Ctv) locus in Poncirus trifoliata L. Raf. *Plant Physiology* 131: 1-11.

YOSHIDA, S., S. MARUYAMA, H. NOZAKI, AND K. SHIRASU. 2010. Horizontal gene transfer by the parasitic plant Striga hermonthica. *Science* 328: 1128.

ZAMBRYSKI, P., AND K. CRAWFORD. 2000. Plasmodesmata: gatekeepers for cell-to-cell transport of developmental signals in plants. *Annual Review of Cell and Developmental Biology* 16: 393-421.

ZEIDAN, M., N. SIKTON, J. COHEN, AND A. GERA. 2001. Improved detection of petunia vein-clearing caulimovirus. *HortScience* 35: 1279-1282.

Tables

TABLE 1. Published records of horizontally transferred genes from bacteria and viruses to angiosperms.

Recipient	Donor	Gene	State	Citation
Several species from the genus *Nicotiana* (Solanaceae) (at least two independent natural transformation events in this genus)	*Agrobacterium rhizogenes*	*rolB*, *rolC*, ORF13, ORF14	Genes active in some species, inactivated in others	Furner et al. (1986),Meyer et al. (1995), Frundt et al. (1998), Aoki and Syono (1999a, 1999b),Intrieri and Buiatti (2001), Suzuki et al. (2002)
N. tabacum, N. tomentosiformis, N. tomentosa, N. kawakamii	Geminivirus-related insertions (*Geminiviridae*)	Significant sequence similarity with the replication origin and the adjacent rep gene	Copies degenerated and rearranged to various extent	Bejarano et al. (1996), Ashby et al. (1997)
Petunia sp. (Solanaceae)	Petunia vein-clearing virus (PVCV) (*Caulimoviridae*)	Full-length viral genome with functional ORFs	Clusters of integrated PVCV DNA; episomal virus can be reconstituted	Richert-Pöggeler and Shepherd (1997)
Musa spp. (Musaceae)	Pararetroviral banana streak virus (*Caulimoviridae*)	Full-length viral genome with functional ORFs	Some copies are not functional; some integrated viral sequence can induce an episomal viral infection	Harper et al. (1999), Ndwora et al. (1999)
Nicotiana tabacum, N. edwardsonii, N. glutinosa, N. sylvestris, N. tomentosiformis	Tobacco vein clearing virus (TVCV)-like insertions (*Caulimovirideae*)	Both full-length virus copies as well as degenerated copies	In some species, episomal virus can be reconstituted	Jakowitsch J et al. (1999), Lockhart et al. (2000), Gregor et al. (2004)
Poncirus trifoliata	Caulimovirus-like	Sequence with high similarity to reverse transcriptase of caulimovirus	Not known	Yang et al. (2003)
Oryza sativa and closely related *Oryza* species	Rice tungro bacilliform virus (RTBV) (*Caulimoviridae*)	Rearranged structures, no intact ORF of the virus	Most likely non-active copies	Kunii et al. (2004)
Solanum tuberosum (Solanaceae)	Pararetrovirus-like sequences	Fragments of viral DNA	No complete copy of pararetrovirus has been recovered yet	Hansen et al. (2005)
Vitis vinifera (Vitaceae)	Potato virus Y (PVY) (*Potyviridae*)	PVY-coat-protein-like cistron	Some potyviral sequences have retained the coat protein ORF which can be expressed	Tanne and Sela (2005)
Dracaena sanderiana (Agavaceae)	Dracaena mottle virus (DrMV)(*Caulimoviridae*)	Genomic southern blot signals with several ORFs of DrMV	Not known	Su et al. (2007)
Lycopersicum esculentum L. hirsutum (Solanaceae)	Pararetrovirus (*Caulimoviridae*)	Full-length viral genome with functional ORFs	Some copies are not functional, some integrated viral sequence can induce an episomal viral infection	Staginnus et al. (2007)
Dahlia variabilis (Asteraceae)	Dahlia mosaic caulimovirus (DMV)(*Caulimoviridae*)	Genomic southern blot signals with ORF1 and ORF4 of DMV	Not known	Pahalawatta et al. (2008)
Vitis vinifera (Vitaceae)	Pararetrovirus-like sequences (*Caulimovirideae*)	Partial ORFs corresponding to reverse transcriptase	Nonactive	Bertsch et al. (2009)

TABLE 2. Published records of putative horizontal gene transfer of non-organellar DNA between angiosperms. Parasitic species are denoted in bold.

Recipient	Donor	Gene	State	Citation
Festuca ovina (*Poaceae*)	*Poa* (*Poaceae*)	*PgiC2*	Both active and degenerated copies are present in the genome	Ghatnekar et al. (2006)
Setaria sp. (*Poaceae*; tr. *Panicoideae*) (direction of the HGT is not clear)	*Oryza* sp. (*Poaceae*; tr. *Bambusoideae*) (direction of the HGT is not clear)	*Mu*-like element	Most likely inactive	Diao et al. (2006)
Several *Oryza* sp. (direction of the HGT is not clear)	*Oryza australiensis* (*Poaceae*) (direction of the HGT is not clear)	*RIRE1* (LTR-retrotransposon family)	Multicopy insertion; some copies are probably active in all species	Roulin et al. (2008)
Lycopersicon sp.	*Arabidopsis thaliana*	*Rider* (Ty1-copia-like retrotransposon)	Active	Cheng et al. (2009)
Oryza sp. (*Poaceae*; tr. *Bambusoideae*)	*Saccharum* sp. (*Poaceae*; tr. *Andropogoneae*)	*Route66* (LTR-retrotransposon)	Active	Roulin et al. (2009)
Striga hermonthica (*Orobanchaceae*)	*Sorghum* (*Poaceae*) or related grass species	Gene encoding a protein of unknown function	Not known	Yoshida (2010)

Published records of horizontally transferred genes located in plant mitochondria. Parasitic species are denoted in bold. [a] adopted from Richardson and Palmer (2006). [b] HGT questioned by Martin (2005) and Goremykin et al. (2009).

GOREMYKIN, V.V., F. SALAMINI, R. VELASCO, AND R. VIOLA. 2009. Mitochondrial DNA of Vitis vinifera and the issue of rampant horizontal gene transfer. *Molecular Biology and Evolution* 26: 99-110.

MARTIN, W. 2005. Lateral gene transfer and other possibilities. *Heredity* 94: 565-566.

RICHARDSON, A. O., AND J. D. PALMER. 2007. Horizontal gene transfer in plants. *Journal of Experimental Botany* 58: 1-9.

Recipient	Donor	Gene(s)	State	Citation
Actinidia	Monocot	*rps2*	Recapture of a gene previously lost to the nucleus	[a,b] Bergthorsson et al. (2003)
Amborella	Eudicot	*atp1*	Duplication of a gene	
Betulaceae	Unclear	*rps11*	Recapture of a gene previously lost to the nucleus	
Caprifoliaceae	Ranunculales	*rps11*	Recapture of a gene previously lost to the nucleus	
Sanguinaria	Monocot	3‘ *rps11*	Creation of a chimeric gene	
Gnetum	*Asterid*	*nad1*B-C	Duplication of a gene	[a]Won and Renner (2003)
Rafflesiaceae	*Tetrastigma* *Vitaceae	*nad1*B-C	Not known	[a]Davis and Wurdack (2004)
Plantago	**Orobanchaceae**	*atp1*	Duplication of a gene	[a]Mower et al. (2004)
Plantago	**Convolvulaceae**	*atp1*	Duplication of a gene	
Apodanthaceae	Fabales	*atp1*	Not known	[a] Nickrent et al. (2004)
Amborella	Angiosperm Angiosperm	*atp4, atp6, atp8, atp9, ccmB, ccmC, ccmF$_{N1}$, cox2 (2×), cox3, nad1, nad2, nad4, nad5, nad7, rpl16, rps19, sdh4*	Duplication of a gene	[a, b] Bergthorsson et al. (2004)
	Moss	*cox2, nad2, nad3, nad4, nad5, nad6, nad7*	Duplication of a gene	
Phaseolus	Angiosperm	cp *pvs-trn*A	Novel gene	[a]Woloszynska et al. (2004)
Ternstroemia	Ericaceae	*atp1*	Not known	[a]Schönenberger et al. (2005)
Bruinsmia	Cyrillaceae	*atp1*	Not known	
Botrychium	**Santalales**	*nad1*B-C, *matR*	Duplication of a gene	[a]Davis et al. (2005)
Rafflesia, Rhizanthes	*Vitis, Tetrastigma*	*atp1*	Not known	Barkman et al. (2007)
Pilostyles	*Pisum, Psorothamnus*	*atp1*	Not known	
Mitrastema	*Fagus, Quercus*	*atp1*	Not known	

A comparison of genomic sequences of organisms listed below against the nuclear genome of *Arabidopsis thaliana* was performed with blastn program and following parameters:

match = 2; mismatch = -3; gap-open = 5; gap-extend = 2

gi|159186452|ref|NC_003064.2| Agrobacterium tumefaciens str. C58 plasmid At, complete sequence

gi|159161952|ref|NC_003065.3| Agrobacterium tumefaciens str. C58 plasmid Ti, complete sequence

gi|159184118|ref|NC_003062.2| Agrobacterium tumefaciens str. C58 chromosome circular, complete sequence

gi|222112705|ref|NC_011994.1| Agrobacterium radiobacter K84 plasmid pAgK84, complete sequence

gi|222108940|ref|NC_011990.1| Agrobacterium radiobacter K84 plasmid pAtK84b, complete sequence

gi|222101962|ref|NC_011987.1| Agrobacterium radiobacter K84 plasmid pAtK84c, complete sequence

gi|10954646|ref|NC_002575.1| Agrobacterium rhizogenes MAFF03-01724 plasmid pRi1724, complete sequence

gi|190404344|ref|NC_010841.1| Agrobacterium rhizogenes plasmid pRi2659, complete sequence

gi|71754380|ref|NC_006277.2| Agrobacterium tumefaciens K84 plasmid pAgK84, complete sequence

gi|190014640|ref|NC_010929.1| Agrobacterium tumefaciens Ti plasmid pTiBo542, complete sequence

gi|10954820|ref|NC_002147.1| Agrobacterium tumefaciens MAFF301001 plasmid pTi-SAKURA, complete sequence

gi|10955016|ref|NC_002377.1| Agrobacterium tumefaciens plasmid Ti, complete sequence

gi|222102329|ref|NC_011986.1| Agrobacterium vitis S4 plasmid pAtS4a, complete sequence

gi|222109117|ref|NC_011991.1| Agrobacterium vitis S4 plasmid pAtS4b, complete sequence

gi|222083145|ref|NC_011984.1| Agrobacterium vitis S4 plasmid pAtS4c, complete sequence

gi|222102412|ref|NC_011981.1| Agrobacterium vitis S4 plasmid pAtS4e, complete sequence

gi|222080117|ref|NC_011982.1| Agrobacterium vitis S4 plasmid pTiS4, complete sequence

gi|218888746|ref|NC_011770.1| Pseudomonas aeruginosa LESB58, complete genome

gi|71733195|ref|NC_005773.3| Pseudomonas syringae pv. phaseolicola 1448A, complete genome

gi|17544719|ref|NC_003295.1| Ralstonia solanacearum GMI1000, complete genome

gi|17548221|ref|NC_003296.1| Ralstonia solanacearum GMI1000 plasmid pGMI1000MP, complete sequence

gi|66766352|ref|NC_007086.1| Xanthomonas campestris pv. campestris str. 8004 chromosome, complete genome

PŘÍLOHA II

Survey of molecular phylogenetics.

Talianova M.

Plant Soil and Environment 53: 413-416, 2007

Survey of molecular phylogenetics

M. Talianová

Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czech Republic

ABSTRACT

Rapidly increasing amount of biological data necessarily requires techniques that would enable to extract the information hidden in the data. Methods of molecular phylogenetics are commonly used tools as well as objects of continuous research within many fields, such as evolutionary biology, systematics, epidemiology, genomics, etc. The evolutionary process not only determines relationships among species, but also allows prediction of structural, physiological and biochemical properties of biomolecules. The article provides the reader with a brief overview of common methods that are currently employed in the field of molecular phylogenetics.

Keywords: evolutionary model; distance-based methods; maximum parsimony; maximum likelihood; Bayesian inference; accuracy of phylogeny

Biological sequences (DNA, RNA and amino acids) are complex sources of genetic variation due to various mechanisms such as local changes in DNA sequences, rearrangements of DNA segments or DNA acquisition by horizontal gene transfer (reviewed in Arber 2000). Thus, the comparative analyses of genes and whole genomes enable an exciting view into evolutionary processes and relationships between genetic materials of different living organisms. The evolutionary process not only determines relationships among species, but also allows prediction of structural, physiological, and biochemical properties (Chambers et al. 2000).

Phylogenetic construction is a hierarchical process

Molecular phylogenetics is a continuously evolving area, using and developing methods that enable to extract necessary information. Most of the techniques used in phylogenetic analyses produce phylogenetic trees (phylogenies), which represent evolutionary histories of compared species. Reconstruction of molecular phylogenetic relationships using DNA, RNA or amino acid sequences is a hierarchical process consisting of four steps: (1) alignment of homological sequences, (2) selection of an appropriate mathematical model describing sequence evolution, (3) application of a suitable tree-building method with regard to the analysed data, and (4) assessment of the quality of the resulting phylogeny and interpretation of obtained results (Steel 2005).

Data and models of sequence evolution

Molecular phylogenetics can utilize various characters, such as genome-level characters (Boore 2006) (e.g. position of mobile genetic elements, genome re-arrangements, gene order position, etc.), but it mostly analyses data in the form of biomolecular sequences (nucleic acids or amino acids). Sequences for phylogenetic study are either generated in laboratory or retrieved from sequence databases and aligned. Correct alignment of sequences is a fundamental prerequisite for phylogenetic relationship reconstruction (Harrison and Langdale 2006). Each of the sequences is a subject

Supported by the Czech Science Foundation, Projects Nos. 204/05/P505 and 521/05/2076, and by the Institutional Research Plan, Project No. AV0Z50040507.

This contribution was presented at the 4th Conference on Methods in Experimental Biology held in 2006 in the Horizont Hotel in Šumava Mts., Czech Republic.

of random (stochastic) influence of very complex evolutionary processes. Although often very simplified, evolutionary processes can be described using mathematical models of evolution. Some models have very simple assumptions, while others are very complex with numerous parameters representing various biologically relevant facts of sequence evolution. Examples of such parameters are branch lengths of trees (interspeciation times and rates of mutation along the branches), parameters associated with the substitution matrix (e.g. transition/transversion bias), or parameters that describe how mutation rates vary across sites in the sequence. The knowledge of the nature of data used in analyses is an important assumption when choosing a model of evolution. The most of the tree-building methods require mathematical model of sequence evolution, to either compute "distances" between sequences (number of differences corrected for backward, parallel or multiple substitutions) or to explicitly evaluate the probabilities of changes between characters (nucleotides or amino acids) in all positions in the sequence. The simplest is the Jukes-Cantor model (Jukes and Cantor 1969) assuming equal frequency of nucleotides and equal substitution rates. More realistic models are HKY model (Hasegawa et al. 1985), General reversible model (GTR) (Rodríguez et al. 1990), Gamma-distributed-rates models (Wakeley 1994, Yang 1994) and Covarion models (Tuffley and Steel 1998). Considering evolution on the protein level, commonly used models are Dayhoff model of protein evolution (Dayhoff et al. 1978), JTT models (Jones et al. 1992), Codon mutation model (Goldman and Yang 1994), VT model (Muller and Vingron 2000), WAG model (Whelan and Goldman 2001) and many others.

The selection and assessment of the most suitable model is a crucial issue in the phylogenetic reconstruction. Various methods exist that enable to statistically test the accuracy of mathematical models. It is possible to perform a comparison of two models using likelihood ratio tests (LRTs) with nested models (i.e. one model is a special case of the second model) (e.g. Huelsenbeck and Crandall 1997), Akaike information criterion (AIC) (Akaike 1974) or Bayesian information criterium (BIC) (Schwarz 1974); otherwise, it is possible to test the overall adequacy of a particular model using parametric bootstrapping (e.g. Whelan et al. 2001) or Bayesian posterior prediction (Huelsenbeck et al. 2001).

Tree-building methods can be classified according to several criteria (Hershkovitz and Leipe 1998). The first way is to define them as either algorithm-based or criterion-based. Algorithm-based methods produce a tree by following a series of steps (e.g. clustering algorithms), while criterion-based methods use an optimality criterion (e.g. the least number of changes in the tree or the topology with a greatest probability of giving rise of analysed data) for comparing alternative phylogenies to one another and deciding, which one fits better. The second group of method-classification is represented by distance-based methods versus character-based methods. Distance-based methods compute pairwise distances according to some measure. Then, the actual data are omitted and the fixed distances are used in the construction of trees. Trees derived using character-based methods are optimised according to the distribution of actual data patterns in relation to a specified character.

Distance-based methods require evolutionary distance (i.e. the number of changes that have occurred along the branches between two sequences) between all pairs of taxa. To obtain relatively unbiased estimate of the evolutionary distance, it is useful to apply a specific evolutionary model that makes assumption about the nature of the evolutionary changes. The examples of distance-based methods used in molecular phylogenetics are the **Least-square method** (Cavalli-Sforza and Edwards 1967, Fitch and Margoliash 1967) or the **Unweighted pair-group method using arithmetic averages – UPGMA** (Sokal and Michener 1958). However, the most popular distance-based technique is the **Neighbor-joining method** (Saitou and Nei 1987) based on agglomerative clustering. Its major strength is the substantial computational speed that makes this method suitable for large datasets; the weakness of this method is the loss of sequence information when converting the data to pairwise distances. It also produces only one tree and thus it is not possible to examine competing hypotheses about the relationship between sequences.

Character-based (discrete) methods operate directly on the aligned sequences rather than on pairwise distances. **Maximum parsimony** (Edwards and Cavalli-Sforza 1963, Fitch 1977) does not require any model of sequence evolution; it just identifies the tree (or trees) that involves the smallest number of mutational changes (i.e. the shortest tree length or fewest evolutionary steps) necessary to explain the differences among the data under investigation. In many cases, MP methods are superior to other techniques because they

are relatively free of assumptions considering nucleotide and amino acid substitution. MP works well when compared sequences are not too divergent, when the rate of nucleotide substitution is relatively constant and the number of nucleotides examined is large. Furthermore, the parsimony analysis is very useful for some types of molecular data (e.g. insertion sequences, insertions/deletions, gene order or short interspersed nuclear elements – SINEs). The typical problem of MP trees is so called "long-branch attraction" (Hendy and Penny 1989) (or similarly "short-branch attraction"). This phenomenon occurs, when rapidly (slowly) evolving sequences are artefactually inferred to be closely related.

Maximum likelihood method (Cavalli-Sforza and Edwards 1967, Felsenstein 1981) requires a stochastic model of sequence evolution over time. The principle of the likelihood is that the explanation, which makes the observed outcome the most likely (i.e. the most probable) to occur, is the one to be preferred. In maximum likelihood, the topology that gives the highest maximum likelihood value is chosen as the final tree. One of the strengths of the maximum likelihood method is the ease with which hypotheses about evolutionary relationships can be formulated. It enables incorporation of complex models to consider biologically important facts of sequence evolution. On the other side, this method is computationally very intensive and thus is not very appropriate for large datasets.

Recently, likelihood-based **Bayesian inference** using Markov chain Monte Carlo technique (Rannala and Yang 1996) has become a popular and very useful method; it has been applied to numerous problems in evolutionary or systematic biology.

Phylogenetic networks (e.g. Maddison 1997, Huson and Bryan 2006, Jin et al. 2006) enable to model evolutionary processes of organisms where non-tree events (reticulations) took part. The reticulations arise due to horizontal gene transfer, hybrid speciation or recombination events, and thus create specific links among organisms.

Accuracy of phylogenetic tree

With the increasing emphasis in biology on reconstruction of phylogenetic trees, questions have arisen as to how confident one should be in a given phylogenetic tree and how the support for phylogenetic trees should be measured. The most commonly used methods are non-parametric bootstrap test (Felsenstein 1985) and jack-knife test (Efron 1982), based on random resampling of the original dataset (Efron 1982). These techniques provide a measure of "confidence" for each clade of an observed tree, based on the proportion of bootstrap trees showing the same branching pattern. Another way of testing the reliability of phylogeny is parametric Bayesian inference (reviewed in Huelsenbeck et al. 2001) where the parameters such as the tree topology, branch lengths, or substitution parameters, are assessed by posterior probabilities.

However, when assessing accuracy of resulting phylogeny, one should be cautious when interpreting the results. Besides relying on test values, various biologically relevant facts causing artefactual relationships in the phylogeny (e.g. bad experiment design, characteristics of the data, sources of homoplasy – parallelism, convergence, horizontal gene transfer) should be considered.

Implementation of phylogenetic methods

On the website http://evolution.genetics.washington.edu/phylip/software.html#methods is a comprehensive overview of various phylogenetic packages and programs. These are arranged according to different criteria, some of them are free, some commercial.

REFERENCES

Akaike H. (1974): A new look at the statistical model identification. IEEE T. Automat. Contr., *19*: 716–723.

Arber W. (2000): Genetic variation: molecular mechanisms and impact on microbial evolution. FEMS Microbiol. Rev., *24*: 1–7.

Boore J.L. (2006): The use of genome-level characters for phylogenetic reconstruction. Trends Ecol. Evol., *21*: 439–446.

Cavalli-Sforza L.L., Edwards A.W.F. (1967): Phylogenetic analysis: Models and estimation procedures. Am. J. Hum. Genet., *19*: 233–257.

Chambers J.K., Macdonald L.E., Sarau H.M., Ames R.S., Freeman K., Foley J.J., Zhu Y., McLaughlin M.M., Murdock P., McMillan L., Trill J., Swift A., Aiyar N., Taylor P., Vawter L., Naheed S., Szekeres P., Hervieu G., Scott C., Watson J.M., Murphy A., Duzic E., Klein C., Bergsma D.J., Wilson S., Livi P. (2000): A G protein-coupled receptor for UDP-glucose. J. Biol. Chem., *15*: 10767–10771.

Dayhoff M.O., Schwartz R.M., Orcutt B.C. (1978): A model of evolutionary change in proteins. In: Dayhoff M.O. (eds.): Atlas of Protein Sequences and Structure. National Biomedical Research Foundation, Silver Spring, MD.

Edwards A.W.F., Cavalli-Sforza L.L. (1963): The reconstruction of evolution. Heredity, *18*: 553.

Efron B. (1982): The Jackknife, the Bootstrap and other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia.

Felsenstein J. (1981): Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol., *17*: 368–376.

Felsenstein J. (1985): Confidence limits on phylogenies: An approach using the bootstrap. Evolution, *39*: 783–791.

Fitch W.M. (1977): On the problem of discovering the most parsimonious tree. Am. Nat., *111*: 223–257.

Fitch W.M., Margoliash E. (1967): Construction of phylogenetic trees. Science, *155*: 279–284.

Goldman N., Yang Z. (1994): A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol., *11*: 725–736.

Harisson C.J., Langdale J.A. (2006): A step by step guide to phylogeny reconstruction. Plant J., *45*: 561–572.

Hasegawa M., Kishino H., Yano T. (1985): Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol., *22*: 160–174.

Hendy M.D., Penny D. (1989): A framework for the quantitative study of evolutionary trees. Syst. Zool., *38*: 297–309.

Hershkovitz M.A., Leipe D.D. (1998): Phylogenetic analysis. In: Baxevanis A.D., Ouellette B.F.F. (eds.): Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. Wiley Interscience, NY.

Huelsenbeck J.P., Crandall K.A. (1997): Phylogeny estimation and hypothesis testing using maximum likelihood. Annu. Rev. Ecol. Syst., *28*: 437–466.

Huelsenbeck J.P., Ronquist F., Nielsen R., Bollback J.P. (2001): Bayesian inference of phylogeny and its impact on evolutionary biology. Science, *294*: 2310–2314.

Huson D.H., Bryant D. (2006): Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol., *23*: 254–267.

Jin G., Nakhleh L., Snir S., Tuller T. (2006): Maximum likelihood of phylogenetic networks. Bioinformatics, *22*: 2604–2611.

Jones D.T., Taylor W.R., Thornton J.M. (1992): The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci., *8*: 275–282.

Jukes T.H., Cantor C.R. (1969): Evolution of protein molecules. In: Munro H.N. (eds.): Mammalian Protein Metabolism. Academic, NY.

Maddison W.P. (1997): Gene trees in species trees. Syst. Biol., *46*: 523–536.

Muller T., Vingron M. (2000): Modeling amino acid replacement. J. Comput. Biol., *7*: 761–776.

Rannala B., Yang Z. (1996): Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. J. Mol. Evol., *43*: 304–311.

Rodríguez F., Oliver J.L., Marin A., Medina R. (1990): The general stochastic model of nucleotide substitution. J. Theor. Biol., *142*: 485–501.

Saitou N., Nei M. (1987): The Neighbor-Joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol., *4*: 406–425.

Schwarz G. (1974): Estimating the dimension of a model. Ann. Stat., *6*: 461–464.

Sokal R.R., Michener C.D. (1958): A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull., *28*: 1409–1438.

Steel M. (2005): Should phylogenetic models be trying to "fit an elephant"? Trends Genet., *21*: 307–309.

Tuffley C., Steel M.A. (1998): Modelling the covarion hypothesis of nucleotide substitution. Math. Biosci., *147*: 63–91.

Wakeley J. (1994): Substitution rate variation among sites and the estimation of transition bias. Mol. Biol. Evol., *11*: 436–442.

Whelan S., Goldman N. (2001): A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol., *18*: 691–699.

Whelan S., Lio P., Goldman N. (2001): Molecular phylogenetics: state-of-the-art methods for looking into the past. Trends Genet., *17*: 262–272.

Yang Z. (1994): Estimating the pattern of nucleotide substitution. J. Mol. Evol., *39*: 105–111.

Received on February 27, 2007

Corresponding author:

Mgr. Martina Talianová, Akademie věd České republiky, Biofyzikální ústav, v. v. i., Královopolská 135, 612 65 Brno, Česká republika
phone: + 420 541 517 247, fax: + 420 541 240 500, e-mail: talianka18@ibp.cz

PŘÍLOHA III

Identification and characterization of a bacteria-like sequence in the genome of some species from the plant genus *Silene*.

Talianová M., Žlůvová J., Hobza R., Vyskot B., Janoušek B.

Biologia Plantarum - přijato do tisku

Identification and characterization of a bacteria-like sequence in the genome of some species from the plant genus *Silene*.

M. Talianová *, J. Žlůvová, R. Hobza, B. Vyskot and B. Janoušek

Department of Plant Developmental Genetics, Institute of Biophysics Czech Academy of Sciences, Královopolská 135, Brno 612 65, Czech Republic

*Corresponding author

Email adresses:
talianka18@ibp.cz
janousek@ibp.cz

Abstract

The aim of this work was to characterize a nucleotide sequence MK14 that originated from a plasmid library obtained via DOP-PCR amplification of laser microdissected Y-chromosomes of *Silene latifolia*. This sequence showed significant similarity to parts of two adjoining genes from bacterial representatives of the genus *Ralstonia*. MK14 sequence contains a part of a conserved domain, and phylogenetic analysis based on this region confirmed its relationship to *Ralstonia*-derived sequences. Genomic Southern blot analysis proved the presence of this fragment in the genome of *S. latifolia*. We hypothesize that this insertion is of bacterial origin, and was probably gained via horizontal gene transfer. Moreover, MK14 insertion is shared by some closely related *Silene* species, suggesting an ancient spontaneous transformation by an ancestor of bacteria from the genus *Ralstonia*.

Additional key words: microdissection, phylogenetic analysis, horizontal gene transfer, transformation, sulfate adenylyltransferase, *Ralstonia*

Abbreviations:

DOP-PCR - Degenerate oligonucleotide primed PCR

HGT - Horizontal gene transfer

Acknowledgement:

We would like to thank the Department of Bacteriology, Crop Research Institute in Prague, for bacteriological tests of *Silene latifolia* seeds. This work was supported by Czech Science Foundation (grants no. 521/08/0932, 522/09/0083 and 204/09/H002) and Institutional Research Plans (AV0Z50040507 and AV0Z50040702). MT was supported by FEBS Short Term Fellowship.

Introduction

The genus *Silene* (*Caryophyllaceae*) is an important model for the studies of various evolutionary aspects (*e.g.*, evolution of sexual systems, evolution of sex chromosomes in their early phases), and in ecological studies (Bernasconi *et al*. 2009, Janousek and Mrackova, 2010). However, only since recently genomic resources for *Silene* have become available greatly facilitating genetic and molecular studies in this genus. Chromosome microdissection, microcloning and construction of chromosome-specific genomic libraries are widely used tools in studying the structure of genomes and identification of chromosome-specific markers (*e.g.*, Mariotti *et al.* 2006, Hobza and Vyskot 2007). Chromosome microdissection has been also used to study the structure of X and Y sex-chromosomes in *S. latifolia* (white campion, *Caryophyllaceae*) (Delichère et *al.* 1999, Matsunaga *et al.* 1999, Sugiyama *et al.* 2003, Hobza *et al.* 2004, Hobza *et al.* 2007). Several markers from the Y-chromosome genomic library were isolated and studied in detail (Hobza *et al.* 2006, Kejnovsky *et al.* 2006). The aim of this work was to characterize another sequence named MK14 that originated from the genomic plasmid library derived from the DOP-PCR on the Y-chromosome of *S. latifolia*.

Materials and Methods

Plant material: In this study, several species of the genus *Silene* were used, including dioecious species *S. latifolia* (population MAV, inbred line kindly provided by Dr. S. Matsunaga, University of Tokio; Bc, interpopulation cross, seed material collection of the Institute of Biophysics, Brno; U9xCH, interpopulation cross, seed material collection of the Institute of Biophysics, Brno; MAVxBystrc, interpopulation cross, seed material collection of the Institute of Biophysics, Brno), *S. dioica* (seed material collection of the Institute of Biophysics, Brno), *S. diclinis* (seed material collection of the Institute of Biophysics, Brno), and gynodioecious species *S. vulgaris* (wild type population, Brno). Plants were grown in greenhouse conditions at 22 °C with 16 h of daylight. Axenic plant cultures from surface-sterilized seeds of *S. latifolia* and *S. vulgaris* were grown on the BMS-10 medium (Ye *et al.* 1990) with antibiotic cefotaxime (0.5 mg cm^{-3}) in sterile conditions for 7 weeks in climate chamber at 22 °C with 16 h of the light. The presence of bacteria from the genus *Burkholderia* and *Ralstonia* in seeds of *S. latifolia* was tested by inoculation of homogenized seeds on the nonspecific King's B medium (King *et al.* 1954).

Isolation of nucleic acids: DNA samples from individual plants were isolated using commercial kits for plant DNA isolation (Qiagen, Germany; Macherey-Nagel, Germany). RNA samples were isolated as described by Zluvova *et al.* (2010) from different tissues (leaves, flower buds of various size: 1 mm, 2 mm, 3 mm, > 3 mm) of both male and female individuals of *S. latifolia* and cDNA was synthesized using the procedure described by Zluvova *et al.* (2006). As a positive control, actin primers were used for the amplification both cDNA and genomic DNA as described by Cegan *et al.* (2010).

PCR amplification: Primers were designed along the MK14 fragment (forward primers: MK14F1n 5'-CGA TAT CGG TAC GGT CAA CA-3', MK14F2n 5'-CGT GCT GGA CGT GCA TAC G-3', MK14Fmt 5'-AAC GAT ATC GGT ACG GTC AA-3'; reverse primers: MK14Rn 5'-AAA AGC AAA GGA ATC CAG GC-3', MK14Rmt 5'-ATG ATC CGT GCA TTT TCC TA-3', MK14R1 5'-ATG ATC CGT GCA TTT TCC TAA TGT GGA T-3'). For all primer combinations, PCR amplification was carried out using the PCR mixture according to Michu et al. (2010) and following PCR cycles: 94 °C/3 min, 35 x (94 °C/ 30 s, 55 °C/ 60 s, 72 °C/ 45 s), 72 °C/ 4 min using 20 ng of total genomic DNA of the studied species (*S. latifolia*, *S. diclinis*, *S. dioica*, and *S. vulgaris*). PCR products were cloned into a vector using commercial cloning kits (pGEM-T easy, Promega, USA). Reverse transcription PCR on the cDNA was run with primers MK14F2n and MK14Rmt using the same program as for standard PCR.

Genomic Southern hybridization: For genomic Southern hybridization, 25-30 μg of each genomic DNA sample was digested with HindIII restriction enzyme (New England Biolabs, USA), because this restriction site is absent in the MK14 sequence. Samples were loaded on 0.8 % agarose gel and transferred onto the positively charged nylon membrane (Amersham, USA) by capillary transfer. Hybridization was carried out for 16 h at 65 °C with a probe derived from a part of conserved CysN_NoDQ_III domain amplified from *S. latifolia* showing high similarity to the corresponding domain from the representatives of the genus *Ralstonia*. The probe was radiolabelled by [α-^{32}P]-dCTP using the Prime-It II Random Primer Labeling Kit (Stratagene, USA).

Phylogenetic analyses: Comparison of MK14 sequence against the public NCBI sequence database was performed using the Blast tool with the default settings. Amino acid sequences of the representatives of sulfate adenylyltransferase large subunit (or large subunit of ATP

sulfurylase) were downloaded from the NCBI database creating a dataset of 4990 sequences. Redundant sequences (*i.e.*, sequences that were 100 % identical) were excluded from the dataset in Jalview editor (Waterhouse *et al.* 2009) reducing the dataset to 1923 amino acid sequences. Resulting dataset was aligned together with bacteria-like amino acid sequences from *Silene* species in Mafft (Katoh *et al.* 2002). Due to the large number of sequences, phylogenetic analysis was performed using approximately-maximum-likelihood phylogenetic inference in FastTree - a program that has been shown to be very efficient (in terms of both resolution and computation time) and thus it is very suitable for the analyses of large datasets (Price *et al.* 2010). In FastTree, statistical support for branches is tested by Shimodaira-Hasegawa (SH) test which has been shown to be a good and very fast alternative to bootstrapping (Guindon *et al.* 2010). When using the SH test, for a given branch, estimated maximum-likelihood branch is compared to the next two most likely nearest-neighbor interchange (NNI) rearrangements of that branch (Guindon *et al.* 2010). Phylogenetic tree was visualized and edited in Dendroscope (Huson *et al.* 2007)

Results and discussion

To characterize sex chromosome specific markers we have analysed sequences from the Y-chromosome specific plasmid library of DOP-PCR products (Hobza *et al.* 2006). A DNA sequence of our interest named MK14 (291 bp long) resulted from random selection of clones from this library. Blasting MK14 sequence against NCBI database revealed strong similarity ($E = 5e^{-152}$) to two adjoining genes in bacterial species from the genus *Ralstonia* (gram-negative Betaproteobacteria, *Burkholderiaceae*), with no significant hits within plant kingdom. One part (85 bp) of this bacteria-like fragment of *S. latifolia* shows similarity to a part of a gene coding for uroporphyrin-III C-methyltransferase (*nirE*), while the second part (203 bp) revealed similarity to the part of a gene encoding the large subunit of sulfate adenylyltransferase (ATP sulfurylase; *cysN*)(Fig. 1). Within the second part, a part of conserved protein domain corresponding to the CysN_NodQ_III protein domain (named according to CysN protein and nodulation protein Q) has been identified. This domain is functionaly related to the domain III of translation elongation factor Tu (EF_Tu), a GTPase which is essential for GTP hydrolysis (Martemyanov and Gudkov 2000). *CysN* and *nodQ* genes have been identified from proteobacteria and few gram-positive bacteria, other eubacteria, archaea and eukaryotes use

different ATP sulfurylase that shows no amino acid similarity to CysN and NodQ (Inagaki *et al.* 2002).

To check whether this bacteria-like sequence does not result from bacterial contamination, we designed a set of primers covering the sequence MK14 and performed PCR, both on tissues of greenhouse-grown plants of *S. latifolia*, *S. diclinis*, *S. dioica* and on the axenic cultures of plants grown from surface sterilized seeds of *S. latifolia* and *S. vulgaris* on the medium containing antibiotic cefotaxime. Moreover, bacteriological tests ruled out the possibility of presence of bacteria from the genus *Ralstonia* in the seeds of *S. latifolia*.

The best results were obtained with the primer combination MK14F2n and MK14Rmt (ranging over the part of conserved CysN_NodQ_III domain) amplifying fragment of expected size (201 bp). Sequencing and comparison of the sequences with the public NCBI databases confirmed that the amplified fragments showed a high similarity to the corresponding DNA sequences of species from the genus *Ralstonia*.

The phylogenetic relationship of the bacteria-like sequences from *Silene* species among all representatives of the sulfate adenylyltransferase (ATP sulfurylase) large subunit retrieved from the NCBI database was examined. Since the dataset was rather large (1928 amino acid sequences), we decided to use approximately-maximum-likelihood phylogenetic inference by FastTree. Phylogenetic tree revealed clustering of bacteria-like sequences derived from the *Silene* species with entries of bacteria from the genus *Ralstonia* with high statistical significance (see a subtree in Fig. 2).

To show genomic organization and to further confirm the presence of the bacteria-like sequences in the genome of *S. latifolia*, we performed Southern blot hybridization on genomic DNA isolated from two different populations. By hybridizing with the probe derived from the CysN_NodQ_III part of *S. latifolia*, signals were obtained for both male and female individuals, and in both populations (one strong signal of 5.6 kb common for all individuals plus four minor intensity signals) (Fig. 3). This suggests that despite the original *Ralstonia*-like fragment MK14 was isolated from the Y-chromosome of *S. latifolia*, it is not Y-specific and most probably there are more than one copy present in the genome. Length polymorphisms were present both within and between populations. Three out of five hybridization signals are common for females and males in both populations. Within U9xCH population, there is a 8.3 kb long fragment that appears to segregate, however no linkage to sexual phenotype was detected in any fragment. Population MAVxBystrc appears to harbor an extra band that is not present in the U9xCH population.

Given the presence of bacteria-like insertion in several closely related *Silene* species (both dioecious and gynodioecious) it is likely that the insertion was already present in the ancestor of these species. Such insertion might have been acquired from some *Ralstonia*-like bacteria. Present day species of this bacterial genus are associated with various ecological niches, such as water, soil, or plant rhizospheres. Some species are phytopathogenic, *e.g., R. solanacearum* (previously known as *Pseudomonas solanacearum*), which is a dangerous pathogen with wide host range, a causal agens of bacterial wilt (Strider *et al.* 1981). Other species (*e.g.*, *R. pickettii*) are associated with infections in humans with attenuated immunity system (Stelzmueller *et al.* 2006).

DNA can be asexually transmitted between more or less distantly related species through a process called horizontal gene transfer (HGT) (Keeling and Palmer 2008). However, it has been thought that higher eukaryotes only seldom take part in this process (Kurland *et al.* 2003). As the amount of sequences is increasing, new cases of HGT including multicellular eukaryotes are recorded (*e.g.*, Keeling and Palmer 2008, Keeling 2009). In plants, numerous cases of foreign DNA uptake have been described including various donors of sequences (*e.g.*, Richardson and Palmer 2007, Bock 2010, Talianova *et al.* submitted). Examples involve genes from the *Ri* plasmid of *Agrobacterium rhizogenes* (*e.g.*, Furner *et al.* 1986), insertions of gemini- and pararetroviral sequences (*e.g.*, Bejarano *et al.* 1996, Staginnus *et al.* 2007), and transfer of genes between plant species involving mitochondrial and nuclear DNA (*e.g.*, Richardson and Palmer 2007, Roulin *et al.* 2009, Yoshida *et al.* 2010). Moreover, there is also a report of HGT from *Ralstonia solanacearum*-like bacterium into potato (*Solanum tuberosum*) genome accompanied by protein domain reshuffling, which has resulted in a mosaic protein conferring resistance to *R. solanacearum* (Feng *et al.* 2003, Talianova *et al.* submitted). Virtually, plants are good candidates for HGT - given the fact that in contrast to animals, plants lack sequestered germline. Thus transformation of any single meristematic cell giving rise to reproductive tissues, or transformation of a cell with a capability to regenerate a novel individual might be sufficient to pass the foreign DNA to further generations.

Several scenarios could explain how the bacterial DNA entered the genome of *Silene*. A spontaneous transformation might have been promoted during infection (*i.e.*, if the ancestral bacterial donor was pathogenic) or symbiosis. An example of well documented HGT is artificial transformation of plants mediated by bacterial pathogens (*A. tumefaciens* and *A. rhizogenes*), and bacterial symbionts (*Sinorhizobium meliloti*, *Mesorhizobium loti*, and *Rhizobium* sp.) once these were equipped with the tumor-inducing plasmid of *A. tumefaciens* (Broothaerts *et al.* 2005). There are yet several other though less likely possibilities for DNA

transfer, such as vector-mediated transfer (*e.g.*, via squash-sucking insects as vehicles for bacteria) or root-mediated absorption of the naked DNA from the soil (Richardson and Palmer 2007).

Mechanisms of HGT are difficult to elucidate, since the cases of HGT are rather detected as ancient events. Some clues can be deduced based on studies of the behaviour of putative donors. A good piece of information is, *e.g.*, available from *Agrobacterium rhizogenes* mediated spontaneous transformation of *Nicotiana* species (*e.g.*, Aoki and Syōno 1999). There is a good knowledge of the action of *Agrobacterium* sp. (*A. rhizogenes* and *A. tumefaciens*) leading to plant transformation promoted by the presence of specialized plasmid. Interestingly, many other bacterial species, including species from the genus *Ralstonia,* have been shown to possess specific kind of mobile DNA called biphenyl catabolic transposon Tn4371 (Toussaint *et al.* 2003, Ryan *et al.* 2009). The region of this integrative conjugative element was found to contain several plasmid-related genes (involved in plasmid replication or partition) together with a cluster of genes corresponding to the type IV secretion system (T4SS). T4SS complexes are associated with the pathogenesis of various bacteria and are known to be involved in functions related to the delivery of substrate molecules to target cells, including the horizontal DNA transfer to both other bacteria and eukaryotic cells (Backert and Meyer 2006). Together with the ability of bacteria to easily share their genetic information this implies that there can be also other bacterial species able to deliver their DNA into the cells of eukaryotic hosts.

Another question is what happens with the DNA once incorporated into the host genome. It has been hypothesized that unless such insertions confer some adaptive role to the host organisms, they are often subjected to genetic degeneration (Keeling and Palmer 2008). In some cases expression of foreign DNA is regulated by host silencing mechanisms (Hobbs *et al.* 1990, Staginnus *et al.* 2007). To see whether the bacteria-like fragment is expressed in *S. latifolia*, we performed reverse transcription PCR with primers MK14F2n and MK14Rmt on RNA samples isolated from leaves and flower buds. However, we have not detected any transcripts of the studied bacteria-like fragment (data now shown). This might be due to the reasons mentioned above - either the sequence does not provide any advantage to the host genome and undergoes degeneration, or some regulatory processes operate on it to prevent its expression in the plant.

Our work further contributes to an increasing amount of detected cases of HGT involving plants and other higher eukaryotes. These cases suggest that natural barriers preventing an uptake of foreign DNA are probably surpassed at a higher frequency than previously thought.

Our results indicate that also other bacteria than *Agrobacterium* sp. could be able to transfer their genetic material across the borderline of bacterial kingdom.

REFERENCES

Aoki, S., Syōno, K.: Horizontal gene transfer and mutation: Ng*rol* genes in the genome of *Nicotiana glauca*. - Proc. Natl. Acad. Sci. U.S.A. **96**: 13229-13234, 1999.

Backert, S., Meyer, T. F.: Type IV secretion systems and their effectors in bacterial pathogenesis. - Curr. Opin. Microbiol. **9**: 207-217, 2006.

Bejarano, E. R., Khashoggi, A., Witty, M., Lichtenstein, C.: Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. - Proc. Natl. Acad. Sci. U.S.A. **93**: 759-764, 1996.

Bernasconi, G., Antonovics, J., Biere, A., Charlesworth, D., Delph, L. F., Filatov, D., Giraud, T., Hood, M. E., Marais, G. A. B., McCauley, D., Pannell, J. R., Shykoff, J. A., Vyskot, B., Wolfe, L. M., Widmer, A.: *Silene* as a model system in ecology and evolution. - Heredity **103**: 5-14, 2009.

Bock, R.: The give-and-take of DNA: horizontal gene transfer in plants. - Trends Plant Sci. **15**: 11-22, 2010.

Broothaerts, W., Mitchell, H. J., Weir, B., Kaines, S., Smith, L. M., Yang, W., Mayer, J. E., Roa-Rodriguez, C., Jefferson, R. A.: Gene transfer to plants by diverse species of bacteria. - Nature **433**: 629-633, 2005.

Cegan, R., Marais, G., Kubekova, H., Blavet, N., Widmer, A., Vyskot, B., Dolezel, J., Safar, J., Hobza, R.: Structure and evolution of *Apetala3*, a sex-linked gene in *Silene latifolia*. - BMC Plant Biol. **10**: e180, 2010.

Delichère, C., Veuskens, J., Hernould, M., Barbacar, N., Mouras, A., Negrutiu, I., Monéger, F.: Sl*Y1*, the first active gene cloned from a plant Y chromosome, encodes a WD-repeat protein. - EMBO J. **18**: 4169-4179, 1999.

Feng, J., Yuan, F., Gao, Y., Liang, C., Xu, J., Zhang, C., He, L.: A novel antimicrobial protein isolated from potato (*Solanum tuberosum*) shares homology with and acid phosphatase. -

Biochem. J. **376**: 481-487, 2003.

Furner, I. J., Huffman, G. A., Amasino, R. M., Garfinkel, D. J., Gordon, M. P., Nester, E. W.: An *Agrobacterium* transformation in the evolution of the genus *Nicotiana*. - Nature **319**: 422-427, 1986.

Guindon, S., Dufayard, J., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O.: New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. - Syst. Biol. **59**: 307-321, 2010.

Hobbs, S. L., Kpodar, P., DeLong, C. M.: The effect of T-DNA copy number, position and methylation on reporter gene expression in tobacco transformants. - Plant Mol. Biol. **15**: 851-864, 1990.

Hobza, R., Lengerova, M., Cernohorska, H., Rubes, J., Vyskot, B.: FAST-FISH with laser beam microdissected DOP-PCR probe distinguishes the sex chromosomes of *Silene latifolia*. - Chromosome Res. **12**: 245-250, 2004.

Hobza, R., Hrusakova, P., Safar, J., Bartos, J., Janousek, B., Zluvova, J., Michu, E., Dolezel, J., Vyskot, B.: MK17, a specific marker closely linked to the gynoecium suppression region on the Y chromosome in *Silene latifolia*. - Theor. Appl. Genet. **113**: 280-287, 2006.

Hobza, R., Vyskot, B.: Laser microdissection-based analysis of plant sex chromosomes. - Methods Cell Biol. **82**: 433-453, 2007.

Hobza, R., Kejnovsky, E., Vyskot, B., Widmer, A.: The role of chromosomal rearrangements in the evolution of *Silene latifolia* sex chromosomes. - Mol. Genet. Genomics **278**: 633-638, 2007.

Huson, D. H., Richter, D. C., Rausch, C., Dezulian, T., Franz, M., Rupp, R.: Dendroscope: An interactive viewer for large phylogenetic trees. - BMC Bioinformatics **8**: 460, 2007.

Inagaki, Y., Doolittle, W., Baldauf, S. L., Roger, A. J.: Lateral transfer of an *EF-1α* gene: origin and evolution of the large subunit of ATP sulfurylase in eubacteria. - Curr. Biol. **12**:

772-776, 2002.

Janousek, B., Mrackova, M.: Sex chromosomes and sex determination pathway dynamics in plant and animal models. - Biol. J. Linn. Soc. Lond. **100**: 737–752, 2010.

Katoh, K., Misawa, K., Kuma, K., Miyata, T.: MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. - Nucl. Acids Res. **30**: 3059-3066, 2002.

Keeling, P. J.: Functional and ecological impacts of horizontal gene transfer in eukaryotes. - Curr. Opin. Genet. Dev. **19**: 613-619, 2009.

Keeling, P. J., Palmer, J. D.: Horizontal gene transfer in eukaryotic evolution. - Nat. Rev. Genet. **9**: 605-618, 2008.

Kejnovsky, E., Kubat, Z., Hobza, R., Lengerova, M., Sato, S., Tabata, S., Fukui, K., Matsunaga, S., Vyskot, B.: Accumulation of chloroplast DNA sequences on the Y chromosome of *Silene latifolia*. - Genetica **128**: 167-175, 2006.

King, E. O., Ward, M. K., Raney, D. E.: Two simple media for the demonstration of pyocyanin and fluorescin. - J. Lab. Clin. Med. **44**: 301-307, 1954.

Kurland, C. G., Canback, B., Berg, O. G.: Horizontal gene transfer: a critical view. - Proc. Natl. Acad. Sci. U.S.A. **100**: 9658-9662, 2003.

Mariotti, B., Navajas-Pérez, R., Lozano, R., Parker, J. S., de la Herrán, R., Rejón, C. R., Rejón, M. R., Garrido-Ramos, M., Jamilena, M.: Cloning and characterization of dispersed repetitive DNA derived from microdissected sex chromosomes of *Rumex acetosa*. - Genome **49**: 114-121, 2006.

Martemyanov, K. A., Gudkov, A. T.: Domain III of elongation factor G from *Thermus thermophilus* is essential for induction of GTP hydrolysis on the ribosome. - J. Biol. Chem. **275**: 35820-35824, 2000.

Matsunaga, S., Kawano, S., Michimoto, T., Higashiyama, T., Nakao, S., Sakai, A., Kuroiwa,

T.: Semi-automatic laser beam microdissection of the Y chromosome and analysis of Y chromosome DNA in a dioecious plant, *Silene latifolia*. - Plant Cell Physiol. **40**: 60-68, 1999.

Michu, E., Mráčková, M., Vyskot, B., Žlůvová, J.: Reduction of heteroduplex formation in PCR amplification. - Biol. Plant. **54**: 173-176, 2010.

Price, M. N., Dehal, P.S., Arkin, A.P.: FastTree2 - approximately maximum-likelihood trees for large alignments. - PLoS One **5**: e9490, 2010.

Richardson, A. O., Palmer, J. D.: Horizontal gene transfer in plants. - J. Exp. Bot. **58**: 1-9, 2007.

Roulin, A., Piegu, B., Fortune, P. M., Sabot, F., D'Hont, A., Manicacci, D., Panaud, O.: Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon Route66 in Poaceae. - BMC Evol. Biol. **9**: 58, 2009.

Ryan, M., Pembroke, J. T., Adley, C.: Novel Tn4371-ICE like element in *Ralstonia pickettii* and genome mining for comparative elements. - BMC Microbiol. **9**: 242, 2009.

Staginnus, C., Gregor, W., Mette, M. F., Teo, C. H., Borroto-Fernández, E. G., Machado, M. L. D. C., Matzke, M., Schwarzacher, T.: Endogenous pararetroviral sequences in tomato (*Solanum lycopersicum*) and related species. - BMC Plant Biol. **7**: 24, 2007.

Stelzmueller, I., Biebl, M., Wiesmayr, S., Eller, M., Hoeller, E., Fille, M., Weiss, G., Lass-Floerl, C, Bonatti, H.: *Ralstonia pickettii* - innocent bystander or a potential threat? - Clin. Microbiol. Infect. **12**: 99-101, 2006.

Strider, D. L., Jones, R. K., Haygood, R. A.: Southern bacterial wilt of geranium caused by *Pseudomonas solanacearum*. - Plant Dis. **65**: 52-53, 1981.

Sugiyama, R., Kazama, Y., Miyazawa, Y., Matsunaga, S., Kawano, S.: CCLS96.1, a member of a multicopy gene family, may encode a non-coding RNA preferentially transcribed in reproductive organs of *Silene latifolia*. - DNA Res. **10**: 213-220, 2003.

Talianova, M., Janousek, B.: What can we learn from tobacco and other Solanaceae about horizontal DNA transfer. - submitted.

Talianova, M., Vyskot, B., Janousek, B.: Interkingdom protein domain fusion: the case of an antimicrobial protein in potato (*Solanum tuberosum*). - submitted.

Toussaint, A., Merlin, C., Monchy, S., Benotmane, M. A., Leplae, R., Mergeay, M., Springael, D.: The biphenyl- and 4-chlorobiphenyl-catabolic transposon Tn4371, a member of a new family of genomic islands related to IncP and Ti plasmids. - Appl. Environ. Microbiol. **69**: 4837-4845, 2003.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., Barton, G.J.: Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. - Bioinformatics **25**: 1189-1191, 2009.

Ye, D., Installé, P., Ciupercescu, D., Veuskens, J., Wu, Y., Salesses, G., Jacobs, M., Negrutiu, I.: Sex determination in the dioecious *Melandrium*. - Sex. Plant Reprod. **3**: 179-186-186, 1990.

Yoshida, S., Maruyama, S., Nozaki, H., Shirasu, K.: Horizontal gene transfer by the parasitic plant *Striga hermonthica*. - Science **328**: 1128, 2010.

Zluvova, J., Nicolas, M., Berger, A., Negrutiu, I., Monéger, F.: Premature arrest of the male flower meristem precedes sexual dimorphism in the dioecious plant *Silene latifolia*. - Proc. Natl. Acad. Sci. U.S.A. **103**: 18854-18859, 2006.

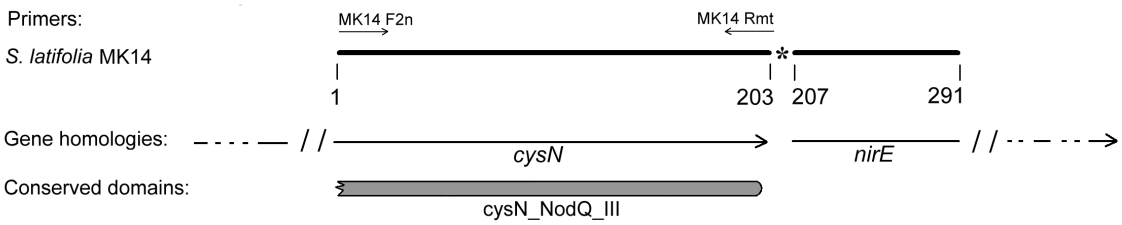
Zluvova, J., Zak, J., Janousek, B., Vyskot, B.: Dioecious *Silene latifolia* plants show sexual dimorphism in the vegetative stage. - BMC Plant Biol. **10**: 208, 2010.

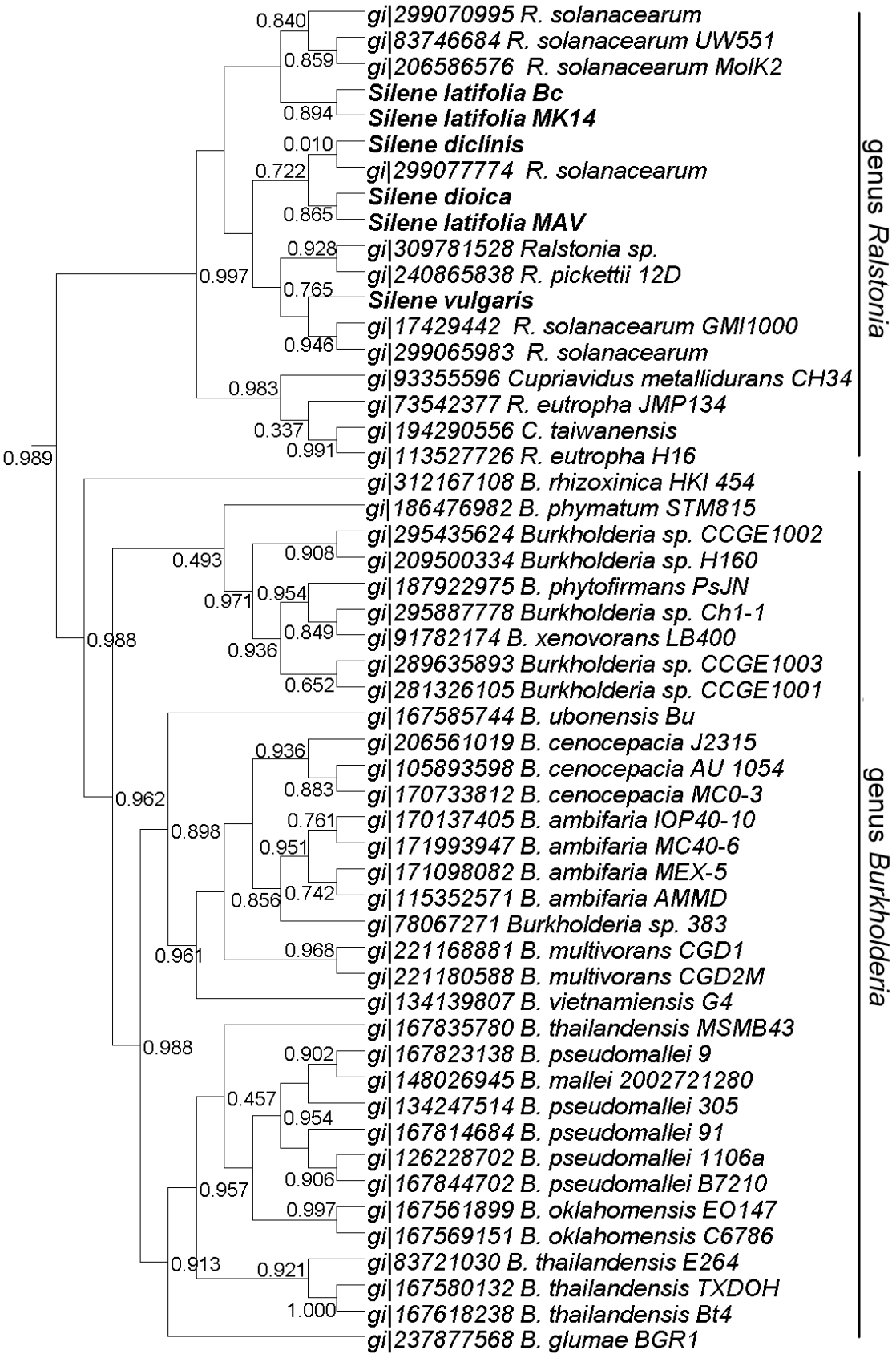
Figure legends

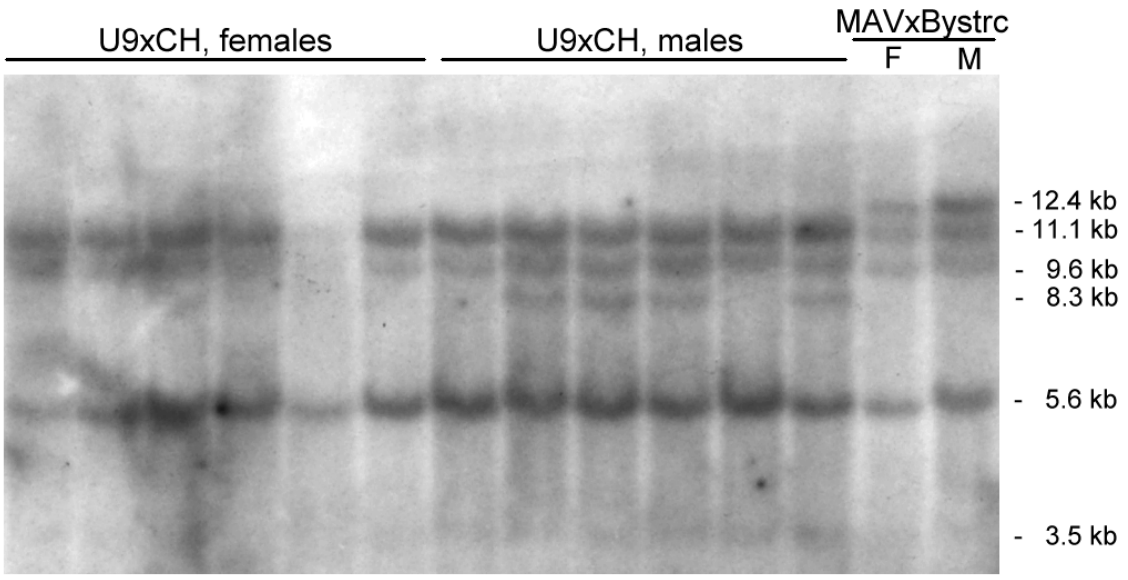
Fig. 1. Graphical representation of *Silene latifolia* MK14 nucleotide sequence along with homologies with corresponding genes *cysN* and *nirE* from *Ralstonia* species, and conservative domain CysN_NodQ_III. Positions of primers MK14F2n and MK14Rmt are denoted. Asterisk represents a termination codon.

Fig. 2. A subtree of FastTree generated approximately-maximum-likelihood phylogeny of sulfate adenylyltransferase large subunit (entire phylogeny available upon request). Sequences from *Silene* species are highlighted in bold. Shimodaira-Hasegawa statistical support values are denoted at nodes (significance is measured as 1-*P*, where *P* is equal to the probability of the null hypothesis - *i.e.*, the reconstructed branch is not significantly more likely than alternative rearrangements).

Fig. 3. Southern hybridization on the HindIII digested genomic DNA from individuals of two different *S. latifolia* populations (six female and six male individuals of the U9xCH population; one female and one male individual of the MAVxBystrc population).


Primers:
MK14 F2n
MK14 Rmt
S. latifolia MK14
1
203
207
291
Gene homologies:
cysN
nirE
Conserved domains:
cysN_NodQ_III


gi|299070995 R. solanacearum
gi|83746684 R. solanacearum UW551
gi|206586576 R. solanacearum MolK2
Silene latifolia Bc
Silene latifolia MK14
Silene diclinis
gi|299077774 R. solanacearum
Silene dioica
Silene latifolia MAV
gi|309781528 Ralstonia sp.
gi|240865838 R. pickettii 12D
Silene vulgaris
gi|17429442 R. solanacearum GMI1000
gi|299065983 R. solanacearum
gi|93355596 Cupriavidus metallidurans CH34
gi|73542377 R. eutropha JMP134
gi|194290556 C. taiwanensis
gi|113527726 R. eutropha H16
gi|312167108 B. rhizoxinica HKI 454
gi|186476982 B. phymatum STM815
gi|295435624 Burkholderia sp. CCGE1002
gi|209500334 Burkholderia sp. H160
gi|187922975 B. phytofirmans PsJN
gi|295887778 Burkholderia sp. Ch1-1
gi|91782174 B. xenovorans LB400
gi|289635893 Burkholderia sp. CCGE1003
gi|281326105 Burkholderia sp. CCGE1001
gi|167585744 B. ubonensis Bu
gi|206561019 B. cenocepacia J2315
gi|105893598 B. cenocepacia AU 1054
gi|170733812 B. cenocepacia MC0-3
gi|170137405 B. ambifaria IOP40-10
gi|171993947 B. ambifaria MC40-6
gi|171098082 B. ambifaria MEX-5
gi|115352571 B. ambifaria AMMD
gi|78067271 Burkholderia sp. 383
gi|221168881 B. multivorans CGD1
gi|221180588 B. multivorans CGD2M
gi|134139807 B. vietnamiensis G4
gi|167835780 B. thailandensis MSMB43
gi|167823138 B. pseudomallei 9
gi|148026945 B. mallei 2002721280
gi|134247514 B. pseudomallei 305
gi|167814684 B. pseudomallei 91
gi|126228702 B. pseudomallei 1106a
gi|167844702 B. pseudomallei B7210
gi|167561899 B. oklahomensis EO147
gi|167569151 B. oklahomensis C6786
gi|83721030 B. thailandensis E264
gi|167580132 B. thailandensis TXDOH
gi|167618238 B. thailandensis Bt4
gi|237877568 B. glumae BGR1
0.840
0.859
0.894
0.010
0.722
0.865
0.928
0.997
0.765
0.946
0.983
0.337
0.991
0.989
0.493
0.908
0.971
0.954
0.849
0.936
0.988
0.652
0.936
0.883
0.962
0.761
0.898
0.951
0.856
0.742
0.961
0.968
0.988
0.902
0.457
0.954
0.906
0.957
0.997
0.913
0.921
1.000
genus Ralstonia
genus Burkholderia


U9xCH, females
U9xCH, males
MAVxBystrc
F
M
- 12.4 kb
- 11.1 kb
- 9.6 kb
- 8.3 kb
- 5.6 kb
- 3.5 kb

PŘÍLOHA IV

Interkingdom protein domain shuffling: the case of an antimicrobial protein in potato (*Solanum tuberosum*).

Talianova M., Vyskot B., Janousek B

odesláno k recenznímu řízení do Plant Systematics and Evolution

Interkingdom protein domain fusion: the case of an antimicrobial protein in potato (*Solanum tuberosum)*.

Martina Talianova [*], Boris Vyskot, Bohuslav Janousek

Department of Plant Developmental Genetics, Institute of Biophysics AS CR, Kralovopolska 135, 612 65, Brno, Czech Republic

* Corresponding Author

Email adresses:

talianka18@ibp.cz

janousek@ibp.cz

Abstract Horizontal gene transfer (HGT) has been thought to play a role in both prokaryotic and eukaryotic evolution. However, the degree to which it shapes eukaryotic genomes is still questionable. The ability to detect and study horizontal gene transfer events is of significant importance for our understanding of its impact on the evolution of eukaryotic genes and genomes. We performed a phylogenetic analysis of a published anti-bacterial protein AP1 from potato (*Solanum tuberosum*). One domain encodes a phosphoesterase that showed high similarity to an acid phosphatase of *Ralstonia solanacearum* and closely related Betaproteobacteria. The second domain encodes an UspA-like domain similar to those present in plants. Our phylogenetic analyses suggest that both domains evolved along different evolutionary pathways until they merged into a single gene. We propose that the phosphoesterase domain was acquired through HGT. Our results support claims in favor of HGT detection at the protein domain level. The case of anti-bacterial protein AP1 in potato highlights the significance of gene fusion/ protein domain fusion as an important feature of horizontal gene transfer, which may greatly contribute to the adaptive abilities of eukaryotic organisms.

Key words horizontal gene transfer, interkingdom gene fusion, novel genes, phylogenetic analysis, antimicrobial protein

Introduction

The enormous progress in molecular techniques and the rapidly expanding amount of genomic data from different organisms has significantly improved our understanding of the evolution of genes with new functions. Several mechanisms are known, alone or in the combination, to give rise to novel genes (reviewed in Long et al. 2003): (1) exon-shuffling, (2) gene duplication, (3) retroposition, (4) mobile elements, (5) gene fusion/ fission, (6) *de novo* origination, and (7) horizontal gene transfer. The degree to which each of these individual mechanisms is understood varies. In this work, we focus on horizontal gene transfer (HGT), a process that results in the transmission of genetic information between different species without sexual reproduction.

HGT has been thought to be an important process in prokaryotic genome evolution, contributing significantly to speciation and adaptation (Doolittle 1999; Keeling and Palmer 2008). The degree to which this process shapes modern eukaryotic genomes remains unclear, and its dynamics and importance are still controversial. Nevertheless, with an increasing amount of sequence data and more sensitive detection methods, new cases of HGT involving eukaryotes are emerging. Despite the many factors that make the detection of HGT difficult, including differential rates of base substitution (Mirkin et al. 2003), lineage-specific gene loss (Huang and Gogarten 2006; Rogers et al. 2007), or loss of phylogenetic signal during the course of evolution, there have been several attempts to estimate the extent of horizontal gene transfer in eukaryotes (e.g. Andersson 2005). Most of these studies have explored whole gene distributions among different organisms, although several authors have claimed that this approach may be too superficial to identify transfer events (Wolf et al. 2000; Choi and Kim 2007; Chan et al. 2009). Chan et al. (2009), in their study of prokaryotic genomes, hypothesized that transferred and recombined regions of DNA might encode intact structural protein domains, which may serve as units of genetic transfer. These authors further demonstrated that within a genome, HGT, accompanied by homologous recombination, can rebuild even the most functionally conservative regions, and thus create genes with mosaic ancestry.

Feng et al. (2003) isolated and characterized an anti-bacterial protein (AP1) (and the corresponding DNA region, NCBI accession AY297449) from the potato (*Solanum tuberosum*) variety MS-42.3. The function of AP1 is to protect plants against two fungal pathogens (*Rhizoctonia solani* and *Alternaria solani)* and several strains of *Ralstonia solanacearum*, a serious bacterial pathogen of crop plants that causes wilt. Feng et al. (2003) noticed that AP1 consisted of two different protein domains. At the C-terminus, they reported sequence similarity to an ATP-binding domain seen in UspA (universal

stress protein A). The ATP-binding domain contains a well-conserved motif mainly related to nucleotide-binding proteins and signal transduction mechanisms (Zarembinski et al. 1998). Feng et al. (2003) further proposed that the N-terminus of AP1 is weakly related to proteins found in plants, showing a high degree of similarity to an acid phosphatase from *Mesorhizobium loti* and from *Burkholderia pseudomallei* (58% and 53% identities respectively).

We retrieved the sequence of AP1 from the NCBI database during our search for putative horizontally transferred sequences from bacteria *Ralstonia solanacearum* to plant genomes. Our aim was to estimate more precisely the evolutionary origin of both domains of AP1. The advantage of this study in comparison with the work of Feng et al. (2003) is that now much more sequencing data is available to create a more complete picture of the evolutionary origin of this interesting protein.

Materials and methods

Data Retrieval and Assembling

To search for homologs of AP1 (both at the nucleotide and amino acid level) in species other than *Solanum tuberosum*, we blasted protein sequence databases and nucleotide databases available at NCBI, as well as in the latest release of the Pfam database (*Solanum tuberosum* AP1 Pfam accession is Q6WBY1). In further analyses, we used the Pfam database as a reference database from which to retrieve sequence data. First, we retrieved amino acid sequences of all representatives of protein domains belonging to phosphoesterase family and Usp family. Second, we extracted sequences for the comparison of AP1 protein domain components with plant versus bacterial family members. For the phosphoesterase family, sequences from all plant entries were assembled together with sequences from *Ralstonia solanacearum* and several other closely related species (*R. metallidurans, R. pickettii, Burkholderia xenovorans, B. graminis, B. glumae, Chromobacterium violaceum*). For the Usp family, due to its wide distribution and high abundance, we chose entries from plant genomes that have been completely sequenced and are present in Pfam (*A. thaliana, M. truncatula, O. sativa, P. trichocarpa, V. vinifera, Z. mays and R. communis*) and bacterial sequences from the same bacterial species used in the search of the phosphoesterase family. Third, a dataset containing AP1 phosphoesterase domain along with phosphoesterase domain sequences from the proteobacterial species was assembled. The datasets corresponding to phosphoesterase domains and to Usp domains, were treated separately.

Further, annotated amino acid sequences of *S. tuberosum* proteins consisting of two or more protein domains were retrieved from Pfam. For each domain, members of the corresponding family were obtained from Pfam, and assembled into an individual dataset. To reduce the amount of data, each dataset was checked for 100 % redundant sequences. Redundancies were removed in Bioedit (Hall 1999) or Jalview (Waterhouse et al. 2009).

Sequence alignments were created using ClustalX (Thompson et al. 1997) and MAFFT (Katoh et al. 2009). Sequences and alignments were edited in SeaView (Galtier et al. 1996) and Geneious (Drummond et al. 2006), multiple alignments were shaded using BOXSHADE (http://www.ch.embnet.org/software/BOX_form.html).

Phylogenetic Analyses

In an effort to maximize efficiency while maintaining reasonable accuracy, we used FastTree (Price et al. 2009), a program that computes approximately-maximum-likelihood phylogenies, to conduct our phylogenetic analyses.We employed the JTT (Jones-Taylor-Thornton)(Jones et al. 1992) model of amino acid evolution with a single rate for each site (the "CAT" approximation) to account for the varying rates of evolution across sites. The reliability (support value) of each split was computed using the Shimodaira-Hasegawa (Shimodaira and Hasegawa 1999) test. For the verification of results, we used PHYML (Guindon and Gascuel 2003) with four categories of gamma distributed substitution rates (all other parameters were estimated from the data), LG model (Le and Gascuel, 2008) and the Shimodaira-Hasegawa test for branch support. We also used MrBayes (Huelsenbeck and Ronquist,) with mixed models, four chains of MCMC run for 1,000,000 generations and with 25% burn-in fraction.

To visualize phylogenetic trees, we used NJplot (Perriere and Gouy 1996) and Dendroscope (Huson et al. 2007).

For calculation of synonymous and non-synonymous substitution rates, we created two pairwise alignments. The first alignment consisted of AP1 phosphoesterase and its closest bacterial homolog *Ralstonia solanacearum* IPO1609 (NCBI accession GI: 206591779). The second alignment consisted of AP1 phosphoesterase and phospholipase C of *Solanum tuberosum* (retrieved from TIGR Solanaceae Genomics Resource). In-frame alignment of protein coding sequences was created using PAL2NAL (Suyama et al. 2006). Synonymous and non-synonymous substitution rates were calculated in the PAML package (Yang 2007) according to the Goldman and Yang (Goldman and Yang 1994) nucleotide substitution model.

Results

Comparison with public databases

The AP1 protein from *Solanum tuberosum* is 343 amino acids long. According to the Pfam records, the phosphoesterase domain has been assigned to the region from 9 – 216 aa, and the UspA domain spans the region from 228 – 306 aa. Low complexity segments are predicted for the regions of 252-263 and 322-334 aa. A comparison against the public databases (NCBI and TIGR Solanaceae Genomics Resource) did not enable us to detect any further homologs of AP1 protein in any

of the organismal kindgoms, nor within the family Solanaceae. A Blast search against the NCBI non-redundant sequences database using the AP1 protein domain components as query produced significant hits to each of the two domains. However, the most statistically significant hits were derived from distinct kingdoms.

At the DNA level, the closest hit was from *Ralstonia solanacearum* (E = 3e-141; Identities = 489/ 624 (78%)) corresponding to the phosphoesterase domain region. The closest hit within plants was from *Ricinus communis* (E = 7e-35; Identities = 179/ 235 (76%)) corresponding to UspA domain. For the phosphoesterase region of AP1, the closest hit among plants was from *Oryza sativa* (E = 1e-16; Identities = 62/ 68), for the UspA region of AP1 there was no significant hit from the bacterial kingdom. At the protein level, the situation was very similar.

Phylogenetic analyses

We conducted an analysis of the evolutionary origin of each protein domain individually. To gain an insight about the position of each domain among all members of corresponding protein domain family, we reconstructed phylogenetic trees based on full sequence alignments (i.e. all sequences available in the Pfam database) (entire phylogenies are available in Supplementary materials, **S1A**, **S1B**). The phylogeny of the phosphoesterase family contains 963 entries across all kingdoms, and the phosphoesterase domain of *Solanum tuberosum* AP1 was positioned deep within a group of bacterial acid phosphatases characteristic of the order *Burkholderiales* (a clade of the tree that groups species from the order *Burkholderiales* contains one sequence from *Chromobacterium violaceum* which belongs to order *Neisseriales*; both orders belong to the class Betaproteobacteria). The phylogenetic tree of the Usp family contains 9002 entries, and places the UspA domain of AP1 deep within plant sequences. Inspired by the results obtained from the full-alignment phylogenies, we wanted to verify the position of each domain within plant family members and family members from selected species of the orders *Burkholderiales* and *Neisseriales* (for the species selection procedure see Materials and Methods). The resulting phylogenetic trees are based on the alignment of 116 amino acid sequences from the phosphoesterase family and 469 amino acid sequences from the Usp family. Due to their large size and for the purpose of clarity, the complexity of both original phylogenetic trees was reduced by collapsing the branches to highlight the position of *Solanum tuberosum* domains among plants and bacteria (for the original entire phylogenetic trees with Pfam accessions, please see Supplementary materials **S2A**, **S2B**). The phylogenies again supported the previous results based

on full alignment trees, positioning the phosphoesterase domain within selected bacterial species (Fig. 1A) and the UspA domain within plant species (Fig. 1B) with high statistical support. Amino acid alignment also suggests that the phosphoesterase domain resembles the bacterial phosphoesterase domain sequences much more than the plant phosphoesterases (Fig. 2A). The same holds for the UspA domain of AP1, which resembles plant sequences more than it resembles bacterial Usp family members (Fig. 2B). Using more robust methods (PHYML and MrBayes), reconstruction of phylogenetic tree of the AP1 phosphoesterase domain and members of phosphoesterase family from Proteobacteria confirmed its close relationship to *Ralstonia*-like species (Fig. 3, data shown only for a part of the tree reconstructed by PHYML). Phylogenetic analyses have revealed few plant entries (3 entries for the phosphoesterase family, 4 entries for the Usp family) that show incongruency by clustering with bacterial family members. In the case of the phosphoesterase family, these entries come from *Populus trichocarpa* (Pfam accession B9PF03; NCBI accession XP_002339596.1), *Oryza sativa* (*indica* cultivar) (NCBI accession AAAA02048760.1) and *Ricinus communis* (Pfam accession B9TF01; NCBI accession XP_002536820.1). The Usp family members showing similarity to bacterial entries come from *Ricinus communis* (Pfam accession B9TK90, B9TM10, B9TPB3; NCBI accession XP_002538659.1, XP_002539279.1, XP_002540082.1), and *Populus trichocarpa* (Pfam accession B9NGQ3; NCBI accession XP_002334439.1). For each of these sequences, the database search revealed an exceptional similarity to sequences from the order *Burkholderiales* species.

We further analyzed 32 proteins (21 proteins of putative nuclear origin, 8 proteins from chloroplasts, and 3 mitochondrial proteins; altogether 79 protein domains) from *S. tuberosum*, which also consist of two or more protein domains (according to the Pfam annotation) (Supplementary materials **S3**). Depending on the quality of sampling (i.e. the availability of sequences in the Pfam database), phylogenetic reconstructions showed that unlike AP1, each of the analyzed protein domains of *S. tuberosum* grouped with plant relatives (or eukaryotic sequences) from the corresponding protein domain family (Supplementary materials **S4.1**-**S4.3**, **S5**, **S6**).

Calculation of divergence times

The divergence time between the phosphoesterase domain of the *Solanum tuberosum* AP1 protein and the closest bacterial homolog from *Ralstonia solanacearum* was estimated by calculating the pairwise synonymous (d_S) and non-

synonymous distance (d_N). We used estimates of substitution rates based on angiosperm actin genes, as calculated by Moniz de Sá and Drouin (1996). The synonymous site distance between the two sequences was 2.7876 substitutions per synonymous site. Assuming a synonymous substitution rate of 6.96×10^{-9} substitutions/site/yr, the date of the divergence is ~ 200 million years ago (MYA). The time divergence based on non-synonymous site changes was ~ 43 MYA, assuming a non-synonymous rate of 0.19×10^{-9} substitutions/site/year and a non-synonymous site distance of 0.1654 substitutions per non-synonymous site. We also attempted to estimate the divergence time between the phosphoesterase domain of the *Solanum tuberosum* AP1 protein and the phosphoesterase domain of phospholipase C of *Solanum tuberosum* as well as for phospholipases C1-C6 in *Arabidopsis thaliana* (the closest homologs from the same family in *Arabidopsis*). This estimation was, however, not possible because of the extreme divergence on the nucleotide level hindering the assembly of reliable alignment necessary for this computation. When in-frame alignment based on protein sequences was applied, we obtained results that do not seem realistic. Based on a synonymous site distance of 55.9209 substitutions per synonymous site, the time divergence would be ~ 4.2 billion years (Gyr) ago, i.e. the approximate date of the origin of life (reviewed by Mattick 2004). When considering a non-synonymous site distance of 0.8967 substitutions per non-synonymous site, the time since the divergence of the two sequences would be ~ 2.35 Gyr ago, a date close to the origin of unicellular eukaryotes. The number of synonymous site changes between AP1 phosphoesterase and phospholipase C is ~62.4 times the number of non-synonymous site changes. Similar estimates were also obtained for the divergence of the phosphoesterase domain of AP1 from the common ancestor of the phospholipases C1-C6 from *Arabidopsis* (d_S = 63.574; d_N=0.7664). These results should not be taken as serious estimates of divergence, but they well illustrate how different the bacteria-like phosphoesterase domain of AP1 is from plant members of the same protein domain family on the nucleotide level.

Discussion

We attempted to trace the evolutionary origin of the anti-bacterial protein AP1 coding gene present in a potato variety resistant to bacterial wilt caused by a serious bacterial plant pathogen *Ralstonia solanacearum* (Feng et al. 2003). Our results confirmed that this protein is composed of two protein domains, a UspA (universal stress protein A) domain and a phosphoesterase domain. Homologs of the UspA domain are ubiquitous in prokaryotes as well as in plants. It is a small

cytoplasmic bacterial protein that is up-regulated when the cell is exposed to various stress factors such as starvation, exposure to toxic chemicals, UV, or osmotic stress, and it has been shown to enhance the rate of cell survival during long exposure to adverse conditions (Nystrom and Neidhardt 1992, 1993, 1996). The phosphoesterase domain (phosphoesterase family) belongs to a group of phosphohydrolases. The phosphoesterase family contains bacterial phospholipase C enzymes as well as eukaryotic acid phosphatases. Phosphoesterases are thought to function mainly in scavenging organic phosphoesters such as nucleotides and sugar phosphates that cannot get through the cytoplasmic membrane (for a review see Rossolini et al.1998). Bacterial phosphatases and phosphohydrolases participate in various cellular processes including virulence and pathogenesis in animals (Terada et al. 1999; Mohapatra et al. 2007). They might also be activated in response to various cellular and environmental stimuli, and they also play an important role in signal transduction through dephosphorylating reactions (Legendre et al. 1993; DeLong 2006). In plants, phosphoesterase domains are also present in proteins related to stress conditions (Kerk et al. 2003).

According to our results, the two domains of AP1 appear to have clearly distinct evolutionary histories. That the phosphoesterase domain in the genome of the rare variety MS-42.3 of *Solanum tuberosum* has a high degree of similarity (> 70 %) to bacterial phosphoesterases along with the absence of both AP1 encoding gene and the phosphoesterase domain in currently available genomic resources of other plant species (if unsure cases are excluded, see below), strongly suggests that the phosphoesterase domain was gained by horizontal gene transfer. As shown by our phylogenetic analyses, the phosphoesterase domain groups with the homologs from the bacterial family *Burkholderiaceae,* suggesting that the putative donor of the bacteria-like domain was related to this bacterial group. Present day species of the family Burkholderiaceae are associated with various ecological niches, such as water, soil, or plant rhizospheres. Some species are phytopathogenic (e.g. *Ralstonia solanacearum*, *Burkholderia cepacia*, *B. glumae*, *B. gladioli*), and other species (e.g. *R. pickettii*, *B. mallei*, *B. cepacia*) are associated with infections in animals and in humans with attenuated immunity (e.g. Strider et al. 1981; Stelzmueller et al. 2006; Stoyanova et al. 2007). Below, we discuss both the horizontal gene transfer hypothesis as well as the alternative processes that could have lead to the origin of the AP1 sequence.

Horizontal gene transfer hypothesis

Horizontal gene transfer (HGT) is a process by which DNA is transmitted asexually between distantly related species (Keeling and Palmer, 2008). While it has been thought that higher eukaryotes are seldomly involved in HGT events (Kurland et al. 2003), with increasing genomic data new cases of HGT involving multicellular eukaryotes have been

recorded (reviewed in Keeling and Palmer 2008; Keeling 2009, Hotopp, 2011). Many of these have been reported in plants and involve donors from a variety of taxonomic groups (reviewed in e.g. Richardson and Palmer 2007; Bock 2010; Talianova and Janousek, submitted). Examples include prokaryote-to-plant HGT (e.g. Furner et al. 1986; Aoki and Syono 1999; Zardoya et al. 2002; Huang and Gogarten 2008), virus-to-plant (e.g. Hull et al. 2000; Murad et al. 2004; Bertsch et al. 2009), plant-to-prokaryote HGT (e.g. Gottig et al. 2009; Pontiroli et al. 2009), or plant-to-plant transfers (e.g. Won and Renner, 2003; Davis and Wurdack 2004; Davis et al. 2005; Stegemann and Bock 2009; Yoshida et al. 2010). In theory, plants are good candidates for HGT. In contrast to animals, their germline is more exposed to external agents, and their genomes appear to be more plastic than those in animals (e.g. Kejnovsky et al. 2009). Thus, transformation of a single meristematic cell that gives rise to reproductive tissues, or transformation of a cell with the ability to regenerate a novel individual might be sufficient to pass the foreign DNA to the next generation.

Several scenarios involving HGT could be proposed as an explanation for the origin of the AP1 coding sequence: the bacterial DNA containing the phosphoesterase domain could have been transferred to the *Solanum tuberosum* genome and then taken part in a recombination or some other rearrangement that resulted in the origin of the chimeric protein, AP1. Alternatively, a fused gene containing both of the domains (a progenitor of AP1) could have also originated in the *Agrobacterium*-like or *Ralstonia*-like species by merging with exogeneous plant DNA. The bacteria could have served as a vector to deliver the fused gene into the plant host. The product of this AP1progenitor gene could have been toxic to a wide range of bacterial vectors, because AP1 has both antibacterial and fungicidal activity (Feng et al. 2003) suggesting that its mechanism of action is relatively nonspecific. Because the phosphoesterase domain and UspA domain are each present in bacteria, it is probably the combination of the two domains in one protein that causes toxicity. A spontaneous transformation involving bacteria could have been promoted during an infection (if the ancestral donor was pathogenic) or during symbiosis. Indeed, both, bacterial plant pathogens (*A. tumefaciens* and *A. rhizogenes*) and symbionts (*Mesorhizobium loti*, *Rhizobium* sp. and *Sinorhizobium meliloti*) have been shown to be capable of transforming plants when equipped with the tumor-inducing plasmid of *A. tumefaciens* (Broothaerts et al. 2005). The possible mechanisms of HGT are difficult to elucidate, since, in most cases, HGTs are rather detected as ancient events. Based on the synonymous substitution rate, we estimated the divergence time of the AP1 phosphoesterase domain to be ~ 200 MYA. However this estimate might be not reliable since the transferred sequences have had to rapidly adapt to a new genomic environment (e.g. different codon preference), and thus have accumulated many synonymous changes. For this reason, a more reliable estimate of divergence time could be the one based on non-synonymous substitution, which suggests ~ 43 MYA since the

divergence of AP1 from the bacterial ancestor sequence. This estimate also corresponds well with the age of the Solanaceae family (~ 40 MYA; Wang et al. 2008), which implies that the bacteria-like phosphoesterase domain should be shared by all Solanaceae species and may not necessarily be unique to *S. tuberosum* which diverged ~ 7.3 MYA from *S. lycopersicum* (Wu and Tanksley 2010). However, because no complete genome of any Solanaceous species is available at the moment, we cannot confirm nor rule out this possibility. On the other hand, both estimates of divergence time might be overestimated, since the extant bacterial species showing the highest similarity to AP1 phosphoesterase are likely to have diverged from the ancient donor.

Nevertheless, certain clues about the mechanisms of spontaneous transfer of the bacteria-like phosphoesterase of AP1 to the plant host can be inferred based on studies of the behaviour of its putative donor. Interestingly, it has been shown that some species from the genus *Ralstonia* (family Burkholderiaceae) possess a biphenyl catabolic transposon Tn4371, which is a specific kind of mobile DNA (Toussaint et al. 2003, Ryan et al. 2009). Several plasmid-related genes and gene clusters corresponding to type IV secretion systems (T4SS) are present on Tn4371. T4SS complexes are associated with the pathogenesis of various bacteria, and are known to be involved in functions related to the delivery of substrate molecules to target cells, including horizontal DNA transfer to both other bacteria and eukaryotic cells (Backert and Meyer 2006). The possible implication is that other bacteria, in addition to *Agrobacterium* species, could also be able to deliver their DNA into the eukaryotic cells. Recently, the HGT of a DNA fragment from *Ralstonia*-like bacteria into the genome of plant species *Silene latifolia* (genus *Caryophyllaceae*) and its close relatives was reported (Talianova et al., in press). When considering adaptive gains of HGT, it seems bacteria have enough to offer. In most cases, however, the adaptive value of DNA transferred to eukaryotes is not obvious. The AP1 protein thus represents a unique example of adaptive gain mediated by HGT.

Alternatives to HGT - paralogy and gene loss

The strong similarity of the AP1 phosphoesterase domain to bacterial homologs, the results of phylogenetic analyses of components of AP1 protein, along with the limited occurrence of AP1 encoding genes and bacteria-like phosphotesterase domains within plant species are highly suggestive of horizontal gene transfer from bacteria. Nevertheless, some attention should be given to alternative evolutionary mechanisms that can produce outcomes similar to those produced by HGT. If we consider the possibility that the AP1 phosphoesterase domain could be of plant origin, then the strong similarity to bacteria would imply strong convergent evolution of ancestral gene, which could be (though not necessarily) preceded by

duplication. Gene duplication followed by differentiation can result in the production of proteins with new functions (Long et al. 2003). According to the classical model, the duplicated gene evolves a new function while the ancestral copy maintains its original function. However, the duplicated gene has to diverge fast enough (which is often accompanied by positive selection) to escape the homogenizing effects, or degeneration. In order to explain the origin of the AP1 phosphoesterase domain using the duplication/divergence hypothesis, we would have to assume that it evolved from an ancestor of bacterial and plant phosphatase domains. A comparison of synonymous and non-synonymous substitution rates between the bacteria-like phosphoesterase of AP1 and the plant-like phosphoesterase domain from phospholipase C from *S. tuberosum* has revealed 62-fold higher rate of synonymous substitutions compared to non-synonymous ones, and their divergence has been dated to 2.3 – 4.0 Gyr ago, corresponding to the origin of the first eukaryotic cells and the origin of life, respectively (reviewed by Mattick 2004). Adaptive evolution could have provoked rapid changes in the gene as it evolved a new function; however, the greatly elevated number of synonymous changes between both domains suggests that high level of divergence between them could not be attributed to adaptive evolution, since adaptive evolution would disproportionately affect non-synonymous changes. As we have already noted in the results, these estimates are probably not very precise, but they nevertheless indicate that evolutionary distance between plant phospholipases C and the AP1 phosphoesterase domain is extremely large. Taking into account that land plants diverged approximately 425 MYA ago (Sanderson 2003), the time since divergence (assuming potential bias in the estimation of divergence time) of both domains appears to be at least long enough for a novel gene (i.e. AP1 phosphoesterase) to be shared by land plants.

A similar argumentation holds for the possibility of convergent evolution without duplication. However, this hypothesis is even less credible, because there must have been extreme selection pressure for a gene/domain to diverge to such extent without being duplicated. Nevertheless, the consequence of proposed hypotheses given the estimated divergence times would be the presence of a bacteria-like domain of AP1 throughout land plants, which is apparently not the case. The absence of such a pattern could be explained by a massive gene loss. Indeed, gene loss has been shown to be a more plausible explanation than HGT in studies of similarities in organellar genes between angiosperms (Goremykin et al. 2009). We found a few plant phosphoesterase sequences (from *P. trichocarpa*, *O. sativa* and *R. communis*) showing incongruent placement among sequences of species from the order *Burkholderiales* and showing significant similarity to the AP1 phosphoesterase domain. However, whether these bacteria-like sequences are truly present in genomes of corresponding species is questionable (two of them have already been removed from the NCBI database: homologs from

P. trichocarpa and *O. sativa*), and thus cannot provide a plausible argument for a gene loss based hypotheses. When all the data are interpreted together, the scenarios based on the convergent evolution and subsequent gene loss in all but two plant species (*S. tuberosum* and maybe *Ricinus communis*) does not appear to be a parsimonious enough explanation for the origin of AP1 gene.

Multidomain proteins and functional variability

The protein domain composition of AP1 suggests that the mechanisms that gave rise to the gene operated in concert, most likely involving horizontal gene transfer and recombination between genomic regions. However, because of lack of the information about the genomic context of the AP1 encoding gene and because a series ancient events was most probably involved, it is difficult to elucidate exact mechanisms of how both domains became coupled together. Several genomic events have been proposed to give rise to such rearrangements ranging from simple point mutations to large-scale chromosomal mutations (reviewed in Moore et al. 2008). On the protein level, rearrangements frequently involve the insertion or deletion of single domains at the N or C terminus. In eukaryotes, a substantial fraction of proteins contain multiple domains (Doolittle 1995; Basu et al. 2008; Koonin et al. 2002). Fusion of domains is thought to have greatly contributed to the formation of various forms of regulation and signaling in the cell (Chinnaiyan et al. 1995; Chen et al. 2002; Koonin et al. 2004; Geisler and Bailly 2007; Itoh et al. 2007). There are debates as to whether the rearrangements of protein domains are random, or are driven by selection for the domain combination based on its function. Support exists for both possibilities, since some domains tend to occur in diverse architectures, while other combinations appear to be driven by random recombination (Apic et al. 2001; Koonin et al. 2002; Vogel et al. 2005). Substantial progress has been made in studying the evolution of protein domains; currently, there is no doubt that due to their specific functionality based on their spatial arrangement, protein domains represent discrete evolutionary units with their own evolutionary history (e.g. Koonin and Wolf 2009; Yang and Bourne 2009). This holds true especially for prokaryotes, since eukaryotic genomes are more stable and large sets of genes tend to evolve congruently. The results of our phylogenetic analysis of individual protein domains of 32 *S. tuberosum* multidomain proteins from the nucleus, chloroplasts and mitochondria is indeed in concordance with the general notion that the major mode of evolution in higher eukaryotes tends to be tree-like (at least at the level of the kingdom), despite the fact that eukaryotic nuclear genes were originally inherited from a common ancestor of bacteria and eukaryotes and that prokaryotic origin is still prominent in chloroplasts and mitochondria.

Cases of interkingdom gene fusions have been reported in bacteria, archaea and some unicellular eukaryotes (e.g. Wolf et al. 2000). In general, during the transfer event, a host cell acquires a foreign external genetic fragment, which is then integrated into the host genome by recombination. The recombination event may also involve the acquisition of parts of other genes (e.g. Cho et al. 1998; Cho and Palmer 1999), individual genes (e.g. Ghatnekar et al. 2006), gene clusters (e.g. Aoki and Syono 1999; Zardoya et al. 2002), or whole plasmids or operons (e.g. Lacroix et al. 2006). Systematic studies (Choi and Kim 2007; Chan et al. 2009) suggest that protein domains should also be regarded as units of horizontal gene transfer. However, in contrast to prokaryotes, the applicability of this concept on eukaryotic genome evolution and the reconstruction of their evolutionary relationships, appears to be minor.

Conclusion

The case of anti-bacterial protein AP1 in potato (*Solanum tuberosum*) is an example of how horizontal gene transfer in connection with the combination of protein domains can be an important source of genes with new functions, which may contribute greatly to the adaptive abilities of eukaryotic organisms. AP1 comprises an exceptionally strong phylogenetic signature, and in other cases we do not expect the phylogenetic signal to be so strong. Although it is difficult to assess the impact of interkingdom protein domain fusion in multicellular eukaryotes, this example again supports the claims that in the assessment of the dynamics and impact of HGT on eukaryotic organisms increased attention should be focused on the evolutionary histories of smaller functional units such as protein domains.

Acknowledgement

This work was supported by the Grant Academy of the Czech Republic (grants no. 521/08/0932, 204/09/H002), by the Czech Ministry of Education (grant no. LC 06 004) and by the Academy of Sciences of the Czech Republic, grants no. AV0Z50040507 and AV0Z50040702.

REFERENCES

Andersson JO (2005) Lateral gene transfer in eukaryotes. Cell Mol Life Sci 62: 1182-1197

Aoki S, Syono K (1999) Horizontal gene transfer and mutation: ngrol genes in the genome of Nicotiana glauca. Proc Natl Acad Sci USA 96: 13229-13234

Apic G, Gough J, Teichmann SA (2001) Domain combinations in archaeal, eubacterial and eukaryotic proteomes. J Mol Biol 310: 311-325

Basu MK, Carmel L, Rogozin IB, Koonin EV (2008) Evolution of protein domain promiscuity in eukaryotes. Genome Res 18: 449-461

Bertsch C, Beuve M, Dolja VV, Wirth M, Pelsy F, Herrbach E, Lemaire O (2009) Retention of the virus-derived sequences in the nuclear genome of grapevine as a potential pathway to virus resistance. Biol Direct 4: 21

Bock R (2010) The give-and-take of DNA: horizontal gene transfer in plants. Trends Plant Sci 15: 11-22

Broothaerts W, Mitchell HJ, Weir B, Kaines S, Smith LM, Yang W, Mayer JE, Roa-Rodriguez C, Jefferson RA (2005) Gene transfer to plants by diverse species of bacteria. Nature 433: 629-633

Davis CC, Wurdack KJ (2004) Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. Science 305: 676-678

Davis CC, Anderson WR, Wurdack KJ (2005) Gene transfer from a parasitic flowering plant to a fern. Proc R Soc Lond B Biol Sci 272: 2237-2242

DeLong A (2006) Switching the flip: protein phosphatase roles in signaling pathways. Curr Opin Plant Biol 9: 470-477

Doolittle RF (1995) The multiplicity of domains in proteins. Annu Rev Biochem 64: 287-314

Doolittle WF (1999) Phylogenetic classification and the universal tree. Science 284: 2124-2129

Drummond AJ, Kearse M, Heled J, Moir R, Thierer T, Ashton B (2006) Geneious v1.2.1, Available from http://www.geneious.com/

Dunning Hotopp JC (2011) Horizontal gene transfer between bacteria and animals. Trends Genet 17: (Epub ahead of print)

Feng J, Yuan F, Gao Y, Liang C, Xu J, Zhang C, He L (2003) A novel antimicrobial protein isolated from potato (Solanum tuberosum) shares homology with an acid phosphatase. Biochem J 376: 481-487

Furner IJ, Huffman GA, Amasino RM, Garfinkel DJ, Gordon MP, Nester EW (1986) An Agrobacterium transformation in the evolution of the genus Nicotiana. Nature 319: 422 – 427

Galtier N, Gouy M, Gautier C (1996) SeaView and Phylo_win, two graphic tools for sequence alignment and molecular phylogeny. Comput Applic Biosci 12: 543-548

Geisler M, Bailly A (2007) Tête-à-tête: the function of FKBPs in plant development. Trends Plant Sci 12: 465-473

Ghatnekar L, Jaarola M, Bengtsson BO (2006) The introgression of a functional nuclear gene from Poa to Festuca ovina. Proc R Soc Lond B Biol Sci 273: 395-399

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11: 725-736

Goremykin VV, Salamini F, Velasco R, Viola R (2009) Mitochondrial DNA of Vitis vinifera and the issue of rampant

horizontal gene transfer. Mol Biol Evol 26: 99-110

Gottig N, Garavaglia BS, Daurelio LD, Valentine A, Gehring C, Orellano EG, Ottado J (2009) Modulating host homeostasis as a strategy in the plant-pathogen arms race. Commun Integr Biol 2: 89-90

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696-704

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41: 95-98

Huang J, Gogarten JP (2006) Ancient horizontal gene transfer can benefit phylogenetic reconstruction. Trends Genet 22: 361-366

Huang J, Gogarten JP (2008) Concerted gene recruitment in early plant evolution. Genome Biol 9: R109

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17: 754-755

Hull R, Harper G, Lockhart B (2000) Viral sequences integrated into plant genomes. Trends Plant Sci 5: 362-365

Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R (2007) Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics 8: 460

Chan CX, Darling AE, Beiko RG, Ragan MA (2009) Are protein domains modules of lateral genetic transfer? PloS ONE 4: e4524

Chen H, Marinissen MJ, Oh S, Chen X, Melnick M, Perrimon N, Gutkind JS, Hou SX (2002) CKA, a novel multidomain protein, regulates the JUN N-terminal kinase signal transduction pathway in Drosophila. Mol Cell Biol 22: 1792-1803

Chinnaiyan AM, O'Rourke K, Tewari M, Dixit VM (1995) FADD, a novel death domain-containing protein, interacts with the death domain of Fas and initiates apoptosis. Cell 81: 505-512

Cho Y, Qiu YL, Kuhlman P, Palmer JD (1998) Explosive invasion of plant mitochondria by a group I intron. Proc Natl Acad Sci USA 95: 14244-14249

Cho Y, Palmer JD (1999) Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial cox1 gene during evolution of the Araceae family. Mol Biol Evol 16: 1155-1165

Choi I, Kim S (2007) Global extent of horizontal gene transfer. Proc Natl Acad Sci USA 104: 4489-4494

Itoh M, Nacher J, Kuma K, Goto S, Kanehisa M (2007) Evolutionary history and functional implications of protein domains and their combinations in eukaryotes. Genome Biol 8: R121

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8: 275-282

Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol 537: 39-64.

Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 9: 605-618

Keeling PJ (2009) Functional and ecological impacts of horizontal gene transfer in eukaryotes. Curr Opin Genet Dev 19: 613-619

Kejnovsky E, Leitch IJ, Leitch AR (2009) Contrasting evolutionary dynamics between angiosperm and mammalian genomes. Trends Ecol Evol 24:572-582

Kerk D, Bulgrien J, Smith DW, Gribskov M (2003) Arabidopsis proteins containing similarity to the universal stress protein domain of bacteria. Plant Physiol 131: 1209-1219

Koonin EV, Wolf YI, Karev WG (2002) The structure of the protein universe and genome evolution. Nature 420: 218-223

Koonin EV, Wolf YI (2009) The fundamental units, processes and patterns of evolution, and the Tree of Life conundrum. Biol Direct 4: 33

Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Rogozin IB, Smirnov S, Sorokin AV, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol 5: R7

Kurland CG, Canback B, Berg OG (2003) Horizontal gene transfer: a critical view. Proc Natl Acad Sci USA 100: 9658-9662

Lacroix B, Tzfira T, Vainstein A, Citovsky V (2006) A case of promiscuity: Agrobacterium's endless hunt for new partners. Trends Genet 22: 29-37

Le S, Gascuel O (2008) An improved general amino-acid replacement matrix. Mol Biol Evol 25: 1307-1320

Legendre L, Yueh YG, Crain R, Haddock N, Heinstein PF, Low PS (1993) Phospholipase C activation during elicitation of the oxidative burst in cultured plant cells. J Biol Chem 268: 24559–24563

Long M, Betrán E, Thornton K, Wang W (2003) The origin of new genes: glimpses from the young and old. Nat Rev Genet 4: 865-875

Mattick JS (2004) RNA regulation: a new genetics? Nat Rev Genet 5:316-323

Mirkin BG, Fenner TI, Galperin MY, Koonin EV (2003) Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. BMC Evol Biol 3: 2

Mohapatra NP, Balagopal A, Soni S, Schlesinger LS, Gunn JS (2007) AcpA is a Francisella acid phosphatase that affects intramacrophage survival and virulence. Infect Immun 75: 390

Moniz de Sá M, Drouin G (1996) Phylogeny and substitution rates of angiosperm actin genes. Mol Biol Evol 13: 1198-1212.

Moore AD, Björklund ÅK, Ekman D, Bornberg-Bauer E, Elofsson A (2008) Arrangements in the modular evolution of proteins. Trends Biochem Sci 33: 441-451

Murad L, Bielawski JP, Matyasek R, Kovarík A, Nichols RA, Leitch AR, Lichtenstein CP (2004) The origin and evolution of geminivirus-related DNA sequences in Nicotiana. Heredity 92: 352-358

Nystrom T, Neidhardt FC (1992) Cloning, mapping and nucleotide sequencing of a gene encoding a universal stress protein in Eschericha coli. Mol Microbiol 6: 3187–3198

Nystrom T, Neidhardt FC (1993) Isolation and properties of a mutant of Escherichia coli with an insertional inactivation of the uspA gene, which encodes a universal stress protein. J Bacteriol 175: 3949–3956

Nystrom T, Neidhardt FC (1996) Effects of overproducing the universal stress protein, UspA, in Escherichia coli K-12. J Bacteriol 178: 927–930

Perriere G, Gouy M (1996) WWW-query: an on-line retrieval system for biological sequence banks. Biochimie 78: 364–369

Pontiroli A, Rizzi A, Simonet P, Daffonchio D, Vogel TM, Monier JM (2009) Visual Evidence of Horizontal Gene Transfer between Plants and Bacteria in the Phytosphere of Transplastomic Tobacco. Appl Environ Microbiol 75: 3314

Price MN, Dehal PS, Arkin AP (2009) FastTree: Computing large minimum-evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26: 1641-1650

Richardson AO, Palmer JD (2007) Horizontal gene transfer in plants. J Exp Bot 58: 1-9

Rogers MB, Watkins RF, Harper JT, Durnford DG, Gray MW, Keeling PJ (2007) A complex and punctate distribution of three eukaryotic genes derived by lateral gene transfer. BMC Evol Biol 7: 89

Rossolini GM, Schippa S, Riccio ML, Berlutti F, Macaskie LE, Thaller MC (1998) Bacterial nonspecific acid phosphohydrolases: physiology, evolution and use as tools in microbial biotechnology. Cell Mol Life Sci 54: 833–850

Ryan M, Pembroke JT, Adley C (2009) Novel Tn4371-ICE like element in Ralstonia pickettii and genome mining for comparative elements. BMC Microbiol 9: 242

Sanderson MJ (2003) Molecular data from 27 proteins do not support a Precambrian origin of land plants. Am J Bot 90: 954-956

Shimodaria H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetics inference. Mol Biol Evol 16: 1114-1116

Stegemann S, Bock R (2009) Exchange of genetic material between cells in plant tissue grafts. Science 324: 649-651

Stelzmueller I, Biebl M, Wiesmayr S, Eller M, Hoeller E, Fille M, Weiss G, Lass-Floerl C, Bonatti H (2006) Ralstonia pickettii – innocent bystander or a potential threat? Clin Microbiol Infect 12: 99-101

Stoyanova M, Pavlina I, Moncheva P, Bogatzevska N (2007) Biodiversity and incidence of Burkholderia species. Biotechnology & Biotechnological Equipment 21: 306 - 310

Strider DL, Jones, RK, Haygood RA (1981) Southern bacterial wilt of geranium caused by Pseudomonas solanacearum. Plant Dis 65: 52-53

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34: W609-W612

Talianová M, Žlůvová J, Hobza R, Vyskot B, Janoušek B (2011) Identification and characterization of a bacteria-like sequence in the genome of some species from the plant genus Silene. Biol Plantarum (in press)

Terada LS, Johansen KA, Nowbar S, Vasil AI, Vasil ML (1999) Pseudomonas aeruginosa hemolytic phospholipase C suppresses neutrophil respiratory burst activity. Infect Immun 67: 2371-2376

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876-4882

Toussaint A, Merlin C, Monchy S, Benotmane MA, Leplae R, Mergeay M, Springael D (2003) The biphenyl- and 4-chlorobiphenyl-catabolic transposon Tn4371, a member of a new family of genomic islands related to IncP and Ti plasmids. Appl Environ Microbiol 69: 4837-4845

Vogel C, Teichmann SA, Pereira-Leal J. (2005) The relationship between domain duplication and recombination. J Mol Biol 346: 355-365

Wang Y, Diehl A, Wu F, Vrebalov J, Giovannoni J, Siepel A, Tanksley SD (2008) Sequencing and comparative analysis of a conserved syntenic segment in the Solanaceae. Genetics 80: 391-408

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version2 – a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191

Wolf YI, Kondrashov AS, Koonin EV (2000) Interkingdom gene fusions. Genome Biol 1: RESEARCH0013

Won H, Renner SS (2003) Horizontal gene transfer from flowering plants to Gnetum. Proc Natl Acad Sci USA 100: 10824-10829

Wu F, Tanksley SD (2010) Chromosomal evolution in the plant family Solanaceae. BMC Genomics 11: 182

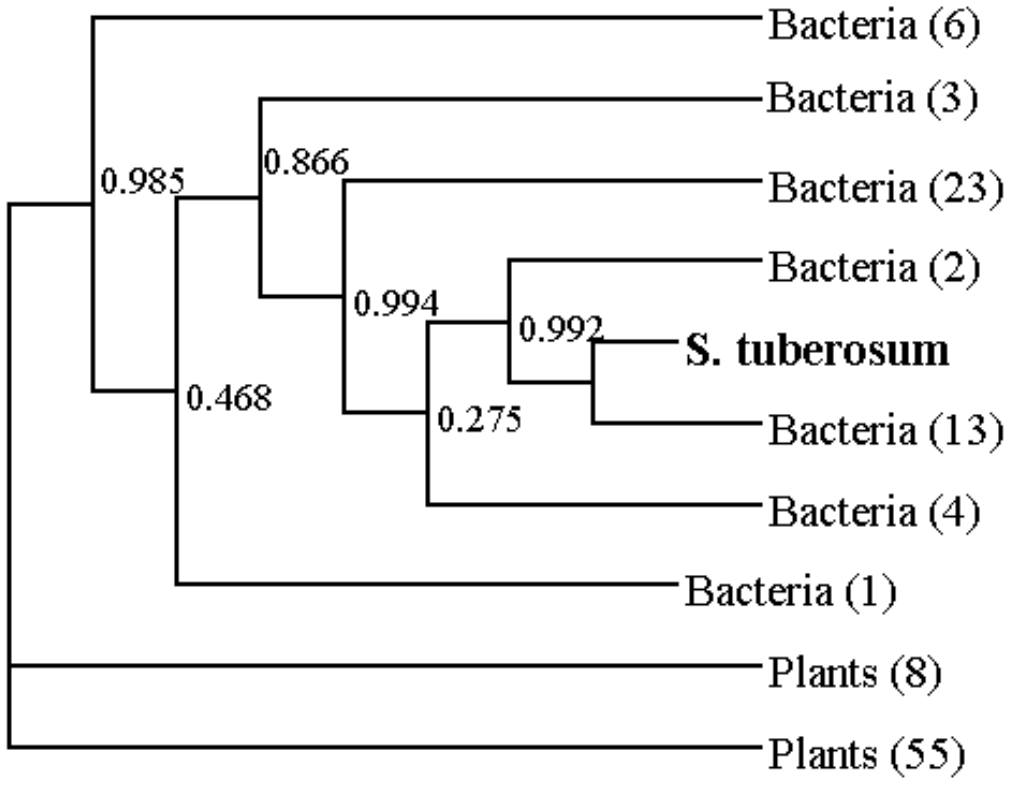
Yang Z (2007) PAML4: a program package for phylogenetic analysis by maximum likelihoo. Mol Biol Evol 24: 1586-1591

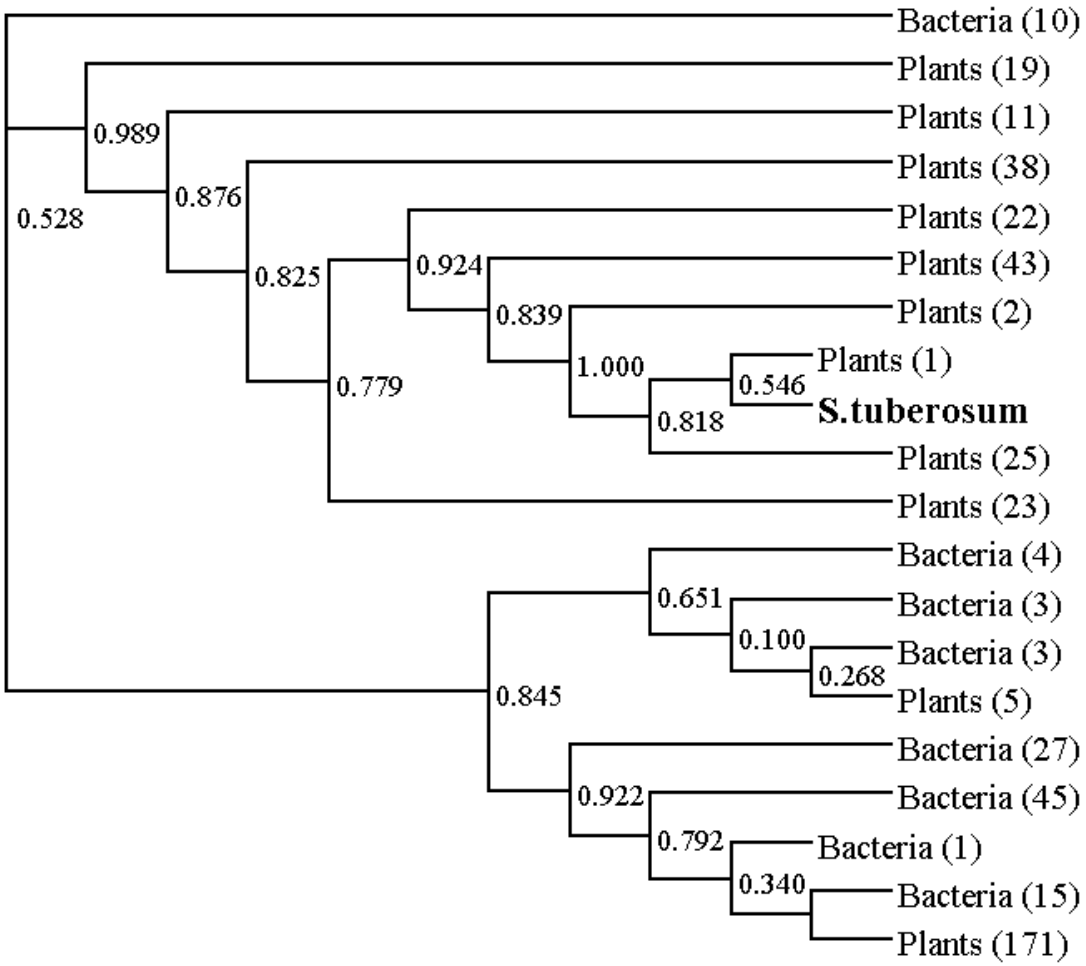
Yang S, Bourne PE (2009) The evolutionary history of protein domains viewed by species phylogeny. PLoS ONE 4: e8378

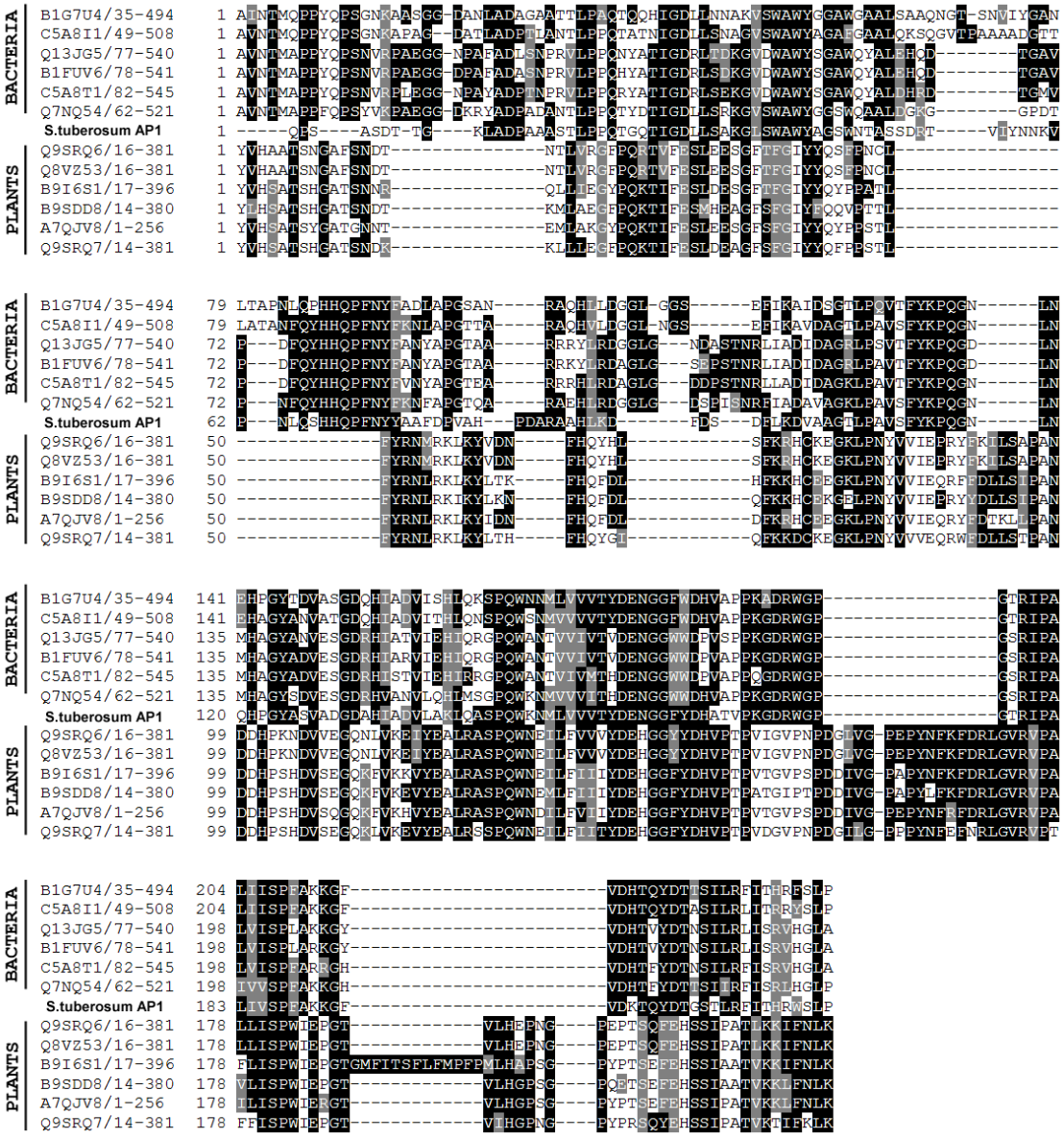
Yoshida S, Maruyama S, Nozaki H, Shirasu K (2010) Horizontal gene transfer by the parasitic plant Striga hermonthica. Science 328:1128.

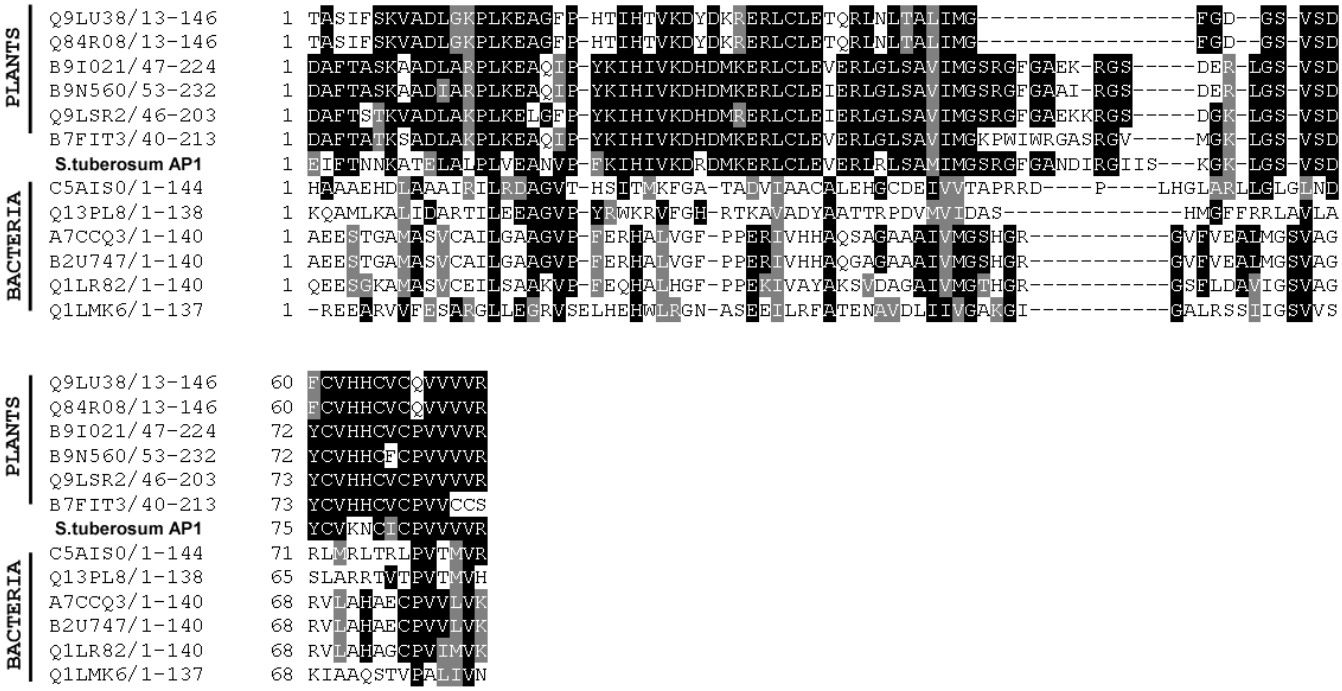
Zardoya R, Ding X, Kitagawa Y, Chrispeels MJ (2002) Origin of plant glycerol transporters by horizontal gene transfer and functional recruitment. Proc Natl Acad Sci USA 99: 14893-14896

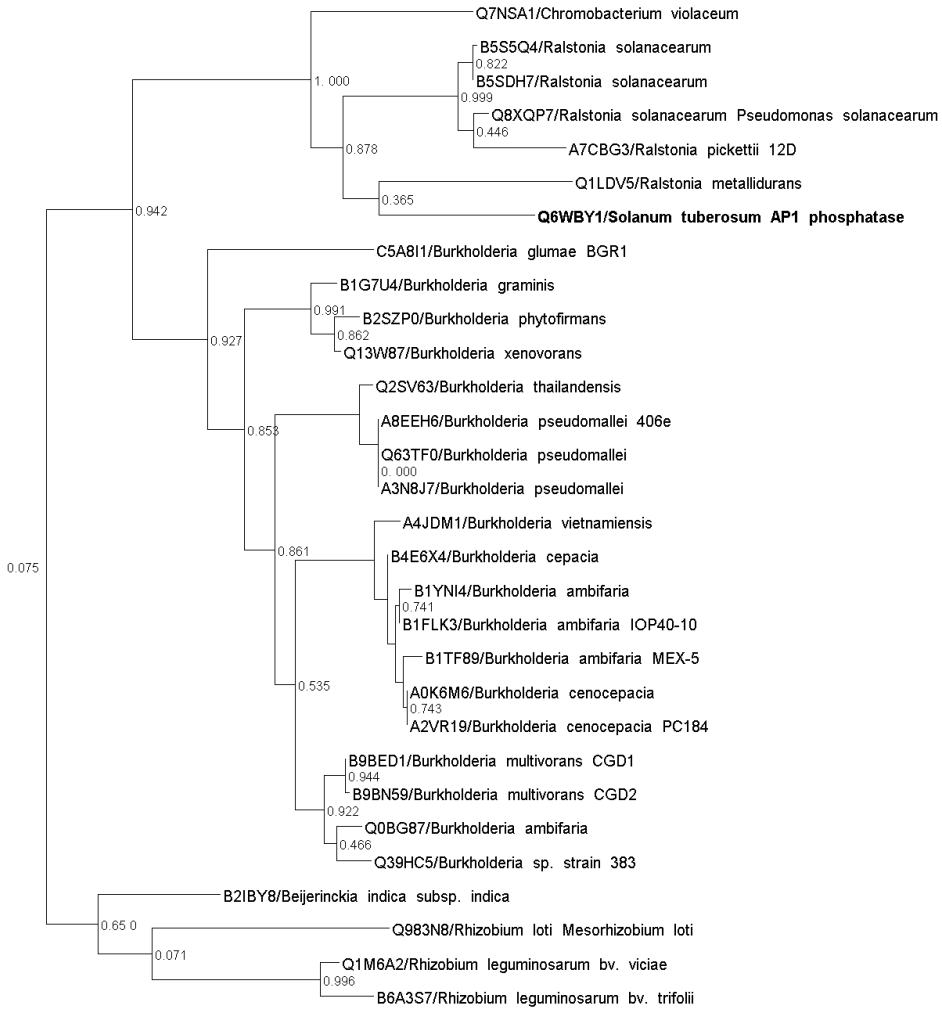
Zarembinski TI, Hung LW, Mueller-Dieckmann HJ, Kim KK, Yokota H, Kim R, Kim SH (1998) Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. Proc Natl Acad Sci USA 95: 15189-15193


Bacteria (6)
Bacteria (3)
Bacteria (23)
Bacteria (2)
S. tuberosum
Bacteria (13)
Bacteria (4)
Bacteria (1)
Plants (8)
Plants (55)
0.985
0.866
0.994
0.992
0.468
0.275


Bacteria (10)
Plants (19)
Plants (11)
Plants (38)
Plants (22)
Plants (43)
Plants (2)
Plants (1)
S.tuberosum
Plants (25)
Plants (23)
Bacteria (4)
Bacteria (3)
Bacteria (3)
Plants (5)
Bacteria (27)
Bacteria (45)
Bacteria (1)
Bacteria (15)
Plants (171)
0.989
0.876
0.528
0.825
0.924
0.839
1.000
0.546
0.818
0.779
0.651
0.100
0.268
0.845
0.922
0.792
0.340


BACTERIA
PLANTS
B1G7U4/35-494
C5A8I1/49-508
Q13JG5/77-540
B1FUV6/78-541
C5A8T1/82-545
Q7NQ54/62-521
S.tuberosum AP1
Q9SRQ6/16-381
Q8VZ53/16-381
B9I6S1/17-396
B9SDD8/14-380
A7QJV8/1-256
Q9SRQ7/14-381


PLANTS
BACTERIA
Q9LU38/13-146
Q84R08/13-146
B9I021/47-224
B9N560/53-232
Q9LSR2/46-203
B7FIT3/40-213
S.tuberosum AP1
C5AIS0/1-144
Q13PL8/1-138
A7CCQ3/1-140
B2U747/1-140
Q1LR82/1-140
Q1LMK6/1-137


Q7NSA1/Chromobacterium violaceum
B5S5Q4/Ralstonia solanacearum
0.822
B5SDH7/Ralstonia solanacearum
1. 000
0.999
Q8XQP7/Ralstonia solanacearum Pseudomonas solanacearum
0.446
A7CBG3/Ralstonia pickettii 12D
0.878
Q1LDV5/Ralstonia metallidurans
0.365
Q6WBY1/Solanum tuberosum AP1 phosphatase
0.942
C5A8I1/Burkholderia glumae BGR1
B1G7U4/Burkholderia graminis
0.991
B2SZP0/Burkholderia phytofirmans
0.862
Q13W87/Burkholderia xenovorans
0.927
Q2SV63/Burkholderia thailandensis
A8EEH6/Burkholderia pseudomallei 406e
0.853
Q63TF0/Burkholderia pseudomallei
0. 000
A3N8J7/Burkholderia pseudomallei
A4JDM1/Burkholderia vietnamiensis
0.861
B4E6X4/Burkholderia cepacia
0.075
B1YNI4/Burkholderia ambifaria
0.741
B1FLK3/Burkholderia ambifaria IOP40-10
B1TF89/Burkholderia ambifaria MEX-5
0.535
A0K6M6/Burkholderia cenocepacia
0.743
A2VR19/Burkholderia cenocepacia PC184
B9BED1/Burkholderia multivorans CGD1
0.944
B9BN59/Burkholderia multivorans CGD2
0.922
Q0BG87/Burkholderia ambifaria
0.466
Q39HC5/Burkholderia sp. strain 383
B2IBY8/Beijerinckia indica subsp. indica
0.65 0
Q983N8/Rhizobium loti Mesorhizobium loti
0.071
Q1M6A2/Rhizobium leguminosarum bv. viciae
0.996
B6A3S7/Rhizobium leguminosarum bv. trifolii

PŘÍLOHA V

Abstrakta z konferencí

ČESKÁ
SPOLEČNOST
EXPERIMENTÁLNÍ
BIOLOGIE ROSTLIN

bulletin

České společnosti
experimentální biologie rostlin
a
Fyziologické sekce
Slovenské botanické společnosti

podzim 2006
4. metodické dny
abstrakta a články

P Přednášky

SURVEY OF MOLECULAR PHYLOGENETICS

MARTINA TALIANOVÁ

Laboratory of Plant Developmental Genetics, Institute of Biophysics, Czech Academy of Sciences, Kralovopolska 135, CZ-612 65 Brno, Czech Republic
E-mail: talianka18@ibp.cz, tel.: +420 541 517 247, fax: +420 541 240 500

Key words: Evolutionary model; Distance-based methods; Maximum parsimony; Maximum likelihood; Bayesian inferrence; Accuracy of phylogeny.

Introduction

Biological sequences (DNA, RNA and amino acids) are complex source of genetic variation due to various mechanisms such as local changes in DNA sequences, rearrangements of DNA segments or DNA acquisition by horizontal gene transfer (reviewed in Arber, 2000). Thus, the comparative analyses of genes and whole genomes enable an exciting view into evolutionary processes and relationships between genetic material of different living organisms. The evolutionary process does not only determine relationships among species, but also allows prediction of structural, physiological, and biochemical properties (Chambers *et al.*, 2000).

Phylogenetic construction is a hierarchical process

Molecular phylogenetics is a continuously evolving area, using and developing methods which enable to extract such information. Most of the techniques used in phylogenetic analyses produce phylogenetic trees (phylogenies) which represent evolutionary histories of compared species. Reconstruction of molecular phylogenetic relationships using DNA, RNA or amino acid sequences is a hierarchical process consisting of four steps: 1) alignment of homological sequences, 2) selection of an appropriate mathematical model describing sequence evolution, 3) application of a suitable tree-building method with regard to the analysed data and 4) assessment of the quality of the resulting phylogeny and interpretation of obtained results (Steel, 2005).

Data and models of sequence evolution

Sequences for phylogenetic study are either generated in laboratory or retrieved from sequence databases and aligned. Correct alignment of sequences is a fundamental prerequisite for phylogenetic relationship reconstruction (Harrison and Langdale, 2006). Each of the sequence is a subject of random (stochastic) influence of very complex evolutionary processes. Although often very simplified, evolutionary processes can be described using mathematical models of evolution. Some models have very simple assumptions, while others are very complex with numerous parameters representing various biologically relevant facts of sequence evolution. Examples of such parameters are branch lengths of the tree (interspeciation times and rates of mutation along the branches), parameters associated with the substitution matrix (e. g., transition/ transversion bias), or parameters that describe how mutation rates vary across sites in the sequence. The knowledge of the nature of data used in analysis is an important assumption when choosing a model of evolution. The most of the tree-building methods require mathematical model of sequence evolution, to either compute ‚distances' between sequences (number of differences corrected for backward, parallel or multiple substitutions) or to explicitly evaluate the probabilities of changes between characters (nucleotides or amino acids) in all positions in the sequence. The simplest model is Jukes-Cantor (Jukes and Cantor, 1969) model assuming equal frequency of nucleotides and equal substitution rates. More realistic models are HKY model (Hasegawa *et al.*, 1985), General reversible model (REV) (Rodríguez *et al.*, 1990), Gamma-distributed-rates models (Wakeley, 1993; Yang, 1994) and Covarion models (Tuffley and Steel, 1998). Considering evolution on the protein level, commonly used models are Codon mutation model (Goldman and Yang, 1994), Dayhoff model of protein evolution (Dayhoff *et al.*, 1978), and many others.

Tree building methods

Tree-building methods can be classified according to several criteria (Hershkovitz and Leipe, 1998). The first way is to define them as either algorithm-based or criterion-based. Algorithm-based methods produce a tree by following a series of steps (e. g., clustering algorithms), while criterion-based methods use an optimality criterion (e. g., the least number of changes in the tree or the topology with a greatest probability of giving rise of analysed data)

for comparing alternative phylogenies to one another and deciding, which one fits better. The second group of method-classification is represented by distance-based methods versus character-based methods. Distance-based methods compute pairwise distances according to some measure. Then, the actual data are omitted and the fixed distances are used in the construction of trees. Trees derived using character-based methods have been optimized according to the distribution of actual data patterns in relation to a specified character.

Distance-based methods require evolutionary distance (e. i., the number of changes that have occurred along the branches between two sequences) between all pairs of taxa. To obtain relatively unbiased estimate of the evolutionary distance, it is useful to apply a specific evolutionary model that makes assumption about the nature of the evolutionary changes. The most popular distance-based technique is *Neighbor-joining* (Saitou and Nei, 1987) method based on agglomerative clustering. Its major strength is the substantial computational speed that makes this method suitable for large dataset. The weakness of this method is the loss of sequence information when converting the data to pairwise distances. It also produces only one tree and thus it is not possible to examine competing hypotheses about the relationship between sequences.

Character-based (discrete) methods operate directly on the aligned sequences rather than on pairwise distances. *Maximum parsimony* (Edwards and Cavalli-Sforza, 1963; Fitch, 1977) does not require any model of sequence evolution, it just identifies the tree (or trees) that involves the smallest number of mutational changes (i. e., the shortest tree length or fewest evolutionary steps) necessary to explain the differences among the data under investigation. In many cases, MP methods are superior to other techniques because they are relatively free from assumptions considering nucleotide and amino acid substitution. MP works well when compared sequences are not too divergent, when the rate of nucleotide substitution is relatively constant and the number of nucleotides examined is large. Furthermore, the parsimony analysis is very useful for some types of molecular data (e. g., insertion sequences, insertions/ deletions, gene order or short interspersed nuclear elements - SINEs). The typical problem of MP trees is so called „long-branch attraction" (Hendy and Penny, 1989) (respectively short-branch attraction). This phenomenon occurs, when rapidly (slowly) evolving sequences are artefactually inferred to be closely related.

The *maximum likelihood method* (Cavalli-Sforza and Edwards, 1967; Felsenstein, 1981) method requires a stochastic model of sequence evolution over time. The principle of the likelihood is that the explanation, that makes the observed outcome the most likely (i. e., the most probable) to occur, is one to be preferred. In maximum likelihood, the topology that gives the highest maximum likelihood value is chosen as the final tree. One of the strengths of the maximum likelihood method is the ease with which hypotheses about evolutionary relationships can be formulated. It enables incorporation of complex models to consider biologically important facts of sequence evolution. On the other side, this method is computationally very intensive and thus it is not very appropriate for large datasets.

Since recently, likelihood-based *Bayesian inferrence* using Markov chain Monte Carlo technique (Rannala and Yang, 1996) becomes popular and very useful method which has been applied to numerous problems in evolutionary or systematic biology.

Accuracy of phylogenetic tree

With the increasing emphasis in biology on reconstruction of phylogenetic trees, questions have arisen as to how confident one should be in a given phylogenetic tree and how support for phylogenetic trees should be measured. The most commonly used methods are non-parametric bootstrap test (Felsenstein, 1985) and jack-knife test (Efron, 1982), based on random resampling of the original dataset (Efron, 1982). These techniques provide a measure of „confidence" for each clade of an observed tree, based on the proportion of bootstrap trees showing that same branching pattern. Another way of testing reliability of phylogeny is parametric Bayesian inferrence (reviewed in Huelsenbeck *et al.*, 2001) where parameters such as the tree topology, branch lengths, or substitution parameters, are assessed by posterior probabilities.

However, when assessing accuracy of resulting phylogeny, one might be cautious when interpreting the results. Except of only relying on test values, various biologically relevant facts causing artefactual relationships in the phylogeny (e. g., bad experiment design, characteristics of the data, sources of homoplasy – parallelism, convergence, horizontal gene transfer) should be accounted.

Implementation of phylogenetic methods

On the website http://evolution.genetics.washington.edu/phylip/software.html#methods is a comprehensive overview of various phylogenetic packages and programs. These are arranged according to different criteria, some of them are free, some are commercial.

Centre Régional de Documentation Pédagogique
Marseille – FRANCE

11th Evolutionary Biology Meeting at Marseilles

September 19th, 20th, 21st, 2007

ASSOCIATION POUR L'ETUDE DE L'EVOLUTION BIOLOGIQUE

POSTER 21

Horizontal gene transfer from the bacterial pathogen *Ralstonia solanacearum* to the plant genome of *Silene latifolia*.

Martina Talianova, Jitka Zluvova, Bohuslav Janousek, Roman Hobza, Boris Vyskot

Laboratory of Plant Developmental Genetics, Institute of Biophysics, Kralovopolska 135, 612 65 Brno, Czech republic

talianka18@ibp.cz

Horizontal gene transfer (HGT) is an asexual transmission of genetic information between unrelated species from various organismal kingdoms. Although in prokaryotes it is considered as a universal mechanism of acquiring novel features, in eukaryotes it occurs in much lower frequency, mainly due to the complexity of eukaryotic organisms. To date, the best studied is the case of horizontal gene transfer from plant bacterial pathogens belonging to the genus *Agrobacterium* to their plant hosts. The horizontal gene transfer is mediated through the oncogenic Ti or Ri plasmids causing tumor transformation in infected plants. These bacteria have evolved a natural ability of stable incorporation of their genetic information (a part of the oncogenic plasmid) into the plant nucleus.

We have analysed the DNA sequence *MK14* originated from the microdissected Y chromosome of the dioecious plant *Silene latifolia* (white campion). This sequence showed a significant homology to the genome of bacterium *Ralstonia solanacearum* which is a devastating plant pathogen with a very wide range of hosts and a global distribution. *MK14* is transcribed and this indicates that it is a part of the gene and it may possess an adaptive function which could be the defense against the pathogen. The horizontally transferred fragment occurs in both dioecious and hermafroditic *Silene* species which implies that the horizontal gene transfer event occured before the speciation of *Silene* species. Our experiments show that *MK14* has more copies in the genome of *Silene* species possessing sex chromosomes. These are located on the Y chromosome and on the X chromosome or autosomes.

In further research, we are interested in estimating the boundaries between the fragment coming from *Ralstonia solanacearum* and the genome of *Silene latifolia*. The phylogenetic analyses in the *Silene* genus will facilitate to follow up the dynamics of evolution and diversification of horizontally transferred sequences in related *Silene* species.

This project was supported by the grants of the Grant agency ČR (M.T. was supported by the grants 521/06/0056, J.Z. by grant 204/05/P505, B.J. by grant 521/05/2076). M.T. obtained a fellowship from FEBS.

SMBE 08

Barcelona
5-8 June 2008

Annual Meeting of the Society for Molecular Biology and Evolution

MEETING PROGRAM

SMBE

UNIVERSITAT DE BARCELONA

ABSTRACT

Analysis of the efficacy of selection in Silene species with different breeding systems and its implication for the evolution of dioecy.

Martina Talianova[1], Jitka Zluvova[1], Elleni Michu[1], Allan Forest[2], Esther Kamau[2], Michael Nicholas[3], Deborah Charlesworth[2], Gabriel Marais[4]

1-Department of Plant Developmental Genetics, Institute of Biophysics AS CR, Brno, Czech Republic.
2-Institute of Evolutionary Biology, King's Buildings, University of Edinburgh, Great Britain
3-Department of Ecology and Evolution, Biophore Building, University of Lausanne, Switzerland
4-Laboratoire de Biométrie et Biologie Evolutive (UMR 5558), CNRS, Univ. Lyon 1, Villeurbanne, France

Contact E-mail: talianka18@ibp.cz

Theoretical works suggest that breeding systems considerably affect genome organisation. Inbreeding and asexuality should lead to genome degeneration. This could explain why inbred and asexual species are more prone to extinction. Experimental studies have shown that polymorphism level may differ in plant species with different reproductive systems. To address these questions, we studied the plant genus Silene,(Caryophyllaceae). It comprises species with different mating systems (hermaphrodites, gynodioecious with different levels of inbreeding and obligately outcrossing dioecious species). The aim of our study was to test the effects of breeding systems on Silene genomes,and more specifically to test the idea that selection is less efficient in non-dioecious than in dioecious species. For this purpose, we used phylogenetic approach based on estimation of synonymous and nonsynonymous substitution rates (dN/dS) across the coding regions of nuclear and plastid genes. Preliminary data based on 5 nuclear genes in 12 Silene species showed that, by contrast to what is expected, dioecious species have significantly higher dN/dS than non-dioecious and suggest that dioecy may not be favoured in the long run, which is consistent with phylogenetic distribution of dioecy in plants. We are currently analysing a larger dataset of 14 nuclear and 2 plastid genes.

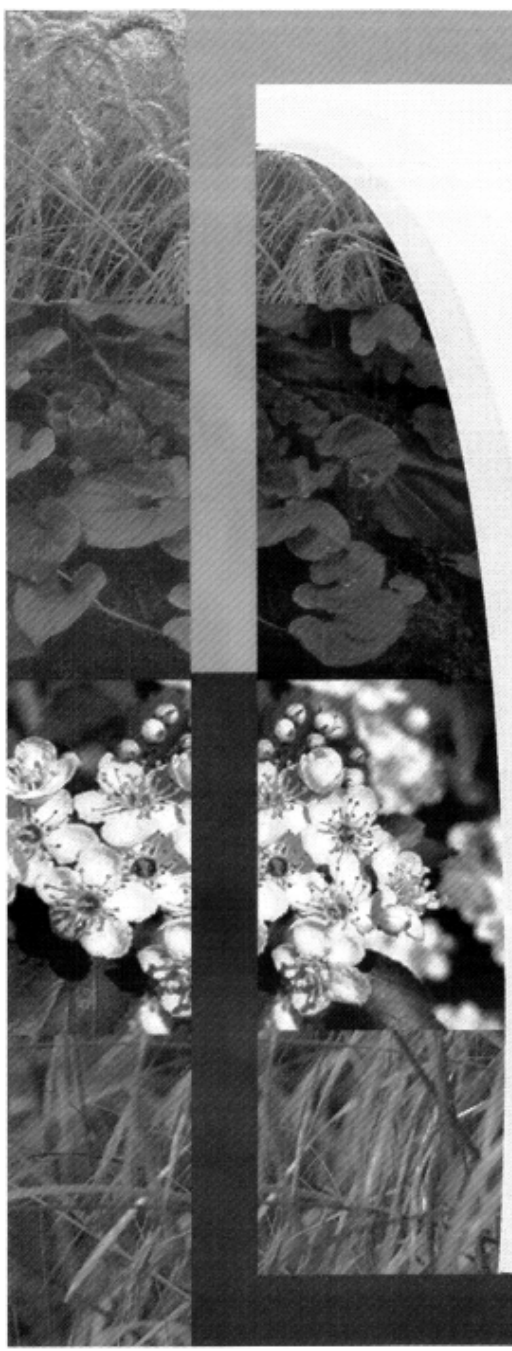
9th IPMB CONGRESS

October 25-30 2009

St. Louis, Missouri, USA

Conference Program

Organized by:
International Society for Plant Molecular Biology
University of Missouri
USDA-Agriculture Research Service



Proceedings of the

9th International Plant Molecular Biology Congress

St. Louis, Missouri USA
October 25-30, 2009

Edited by Perry Gustafson

ISPMB

Abstract CD sponsored by

BASF
The Chemical Company

9th IPMB CONGRESS
October 25-30
2009
St. Louis, Missouri, USA

Click on page to enter proceedings

HORIZONTAL GENE TRANSFER – NEW CASES IN PLANTS

Authors: Martina Talianova; Jitka Zluvova; Roman Hobza; Boris Vyskot; Bohuslav Janousek, Department of Plant Developmental Genetics, Institute of Biophysics AS CR, Czech Republic

Presenter: Ms. Martina Talianova

Horizontal gene transfer (HGT) has been thought to play an important role in the evolution of genomes, especially of the prokaryotic ones, promoting innovation and significantly contributing to speciation and adaptation. Concerning eukaryotes, apart from the transfer of genes from organelles to the nucleus that has been well documented in some eukaryotic lineages, interkingdom gene transmission has been thought to be far less common. Here, we present new cases of HGT between bacteria and plants illustrating some interesting aspects of horizontal gene transmission that may lead to adaptive novelties. One of the cases is an insertion of a sequence named MK14 originated from bacteria Ralstonia solanacearum into the nuclear genome of closely related species from the plant genus Silene. This 330bp long sequence was isolated from the microdissected Y-chromosome of S. latifolia and shows significant similarity to the database sequence of this serious bacterial plant pathogen. Searching the public databases we also found and studied other sequences in plants that seem to originate from the horizontal gene transfer between plants and bacteria. We discuss two main topics, (1) the formation of chimerical proteins promoted by protein domain reshuffling caused by HGT, and (2) interactions between plants and a plant bacterial pathogen that can also be a source of horizontal transfer of genes.

ČESKÁ
SPOLEČNOST
EXPERIMENTÁLNÍ
BIOLOGIE ROSTLIN

bulletin

České společnosti experimentální biologie rostlin a Fyziologické sekce Slovenské botanické společnosti

podzim 2010

12. KONFERENCE EXPERIMENTÁLNÍ BIOLOGIE ROSTLIN

8. KONFERENCE DOKTORANDŮ EXPERIMENTÁLNÍ BIOLOGIE ROSTLIN

sborník abstraktů

P Přednášky

HORIZONTÁLNY PRENOS DNA – ČO NÁM O ŇOM MÔŽU PREZRADIŤ RASTLINY?

MARTINA TALIANOVÁ, BOHUSLAV JANOUŠEK

Oddělení vývojové genetiky rostlin, Biofyzikální ústav AV ČR, Královopolská 135, 612 65 Brno, ČR
E-mail: talianka18@ibp.cz

Horizontálny génový prenos (HGP), tj. prenos genetickej informácie medzi druhmi skrz reprodukčné bariéry, sa považuje za jeden z dôležitých mechanizmov, ktoré sa podieľajú na evolúcií genómov, adaptácií a špeciácií organizmov. U prokaryotických organizmov hrá HGP kľúčovú úlohu – takýmto spôsobom napríklad baktérie veľmi rýchlo získavajú gény rezistencie voči antibiotikám, či rozličné faktory virulencie. U eukaryotických organizmov (zvlášť u mnohobunkových) dochádza k HGP v omnoho menšej miere, čo predovšetkým súvisí s ich komplexitou. Väčšina prípadov HGP zahŕňajúceho mnohobunkové organizmy bola doposiaľ identifikovaná u rastlín. V súčasnosti najlepšie preštudovaný prípad HGP u eukaryot je transformácia rastlín pomocou baktérií z rodu Agrobacterium (A. tumefaciens a A. rhizogenes). V našej práci sa venujeme analýze ďalších prípadov HGP u rastlín, konkrétne u rodu Silene a u zemiaku (Solanum tuberosum). V oboch prípadoch sa jedná o prenos genetického materiálu z bakteriálneho donora blízkeho rodu Ralstonia (Burkholderiaceae, Betaproteobacteria). Zaujímajú nás predovšetkým otázky súvisiace s evolučným významom horizontálneho prenosu genetického materiálu pre rastliny.

Poďakovanie: Táto práca bola podporená grantami GAČR (521/08/0932 a 204/09/H002)

IS ZEATIN CIS-TRANS ISOMERASE A REAL PROTEIN?

TOMÁŠ HLUSKA[1], OLGA RYPAROVÁ[1], KATEŘINA VÁCLAVÍKOVÁ[1,2], JANA KLÁSKOVÁ[1], LUCIE ŠVEHLOVÁ[3], RENÉ LENOBEL[2], MAREK ŠEBELA[3] AND PETR GALUSZKA[1]

[1]Division of Molecular Biology, 2Laboratory of Growth Regulators and 3Department of Biochemistry, Palacký University & Institute of Experimental Botany AS CR, Olomouc, Czech Republic
E-mail: tomas.hluska@upol.cz

Zeatins are isoprenoid cytokinins (CKs) with hydroxyl group at the terminal carbon either in cis- or trans-position. While trans-zeatin is the predominant type of CKs in many plants, cis-zeatin was thought to be present only in scarce amounts with no or low biological activity reflected by its ignorance during cytokinin purifications and term "zeatin" preferably used for trans-isomer. Also the study of metabolism was focused on trans-zeatin with isopentenyladenine discovering their biosynthetic and degradation pathways.
However, recently cis-zeatin is being found to be the predominant cytokinin in an increasing number of plant species. The biosynthesis of cis-zeatin is not known, it is thought that cis-zeatin is released from tRNA after hydroxylation of prenylated adenine, but the data from Arabidopsis tRNA-IPT knock-out lines suggest that there could be even cis-hydroxylated side chain precursor.
The zeatin cis-trans isomerase was described already 17 years ago (Bassil et al. 1993), but the putative protein has not been identified. Existence of this protein is questionable due to several reasons. First, the reaction can run non-enzymaticaly only in the presence of needed "cofactors". Second, the enzymatic reaction has not been shown in planta yet: the experiments with tracer cytokinins showed no conversion and Arabidopsis plants lacking IPT activity were deficient in either cis- or trans-zeatin in accordance to proposed distinctive side chain origin.
Here we will discuss, whether is the zeatin cis-trans isomerase a real protein or is it only non-enzymic artifact. We will show results of purification and partial characterization of the maize zeatin cis-trans isomerase.

Literature:
Bassil NV, Mok DWS and Mok MC (1993) Partial Purification of a cis-trans-Isomerase of Zeatin from Immature Seed of Phaseolus vulgaris L. Plant. Physiol. 102, 867-872