

SHORT REPORT

Open Access

Discovery of novel plastid phenylalanine (*trnF*) pseudogenes defines a distinctive clade in Solanaceae

Péter Poczaï* and Jaakko Hyvönen

Abstract

Background: The plastome of embryophytes is known for its high degree of conservation in size, structure, gene content and linear order of genes. The duplication of entire tRNA genes or their arrangement in a tandem array composed by multiple pseudogene copies is extremely rare in the plastome. Pseudogene repeats of the *trnF* gene have rarely been described from the chloroplast genome of angiosperms.

Findings: We report the discovery of duplicated copies of the original phenylalanine (*trnF_{GAA}*) gene in Solanaceae that are specific to a larger clade within the Solanoideae subfamily. The pseudogene copies are composed of several highly structured motifs that are partial residues or entire parts of the anticodon, T- and D-domains of the original *trnF* gene.

Conclusions: The Pseudosolanoid clade consists of 29 genera and includes many economically important plants such as potato, tomato, eggplant and pepper.

Keywords: Chloroplast DNA (cpDNA); Gene duplications; Phylogeny; Plastome evolution; Tandem repeats; *trnL-trnF*; Solanaceae

Findings

The plastid *trnT-trnF* region has been widely applied to resolve phylogeny of embryophytes (Quandt and Stech 2004; Zhao et al. 2011) and to address various questions of population genetics since the development of universal primers by Taberlet et al. (1991). This marker is located in the large single copy region of the chloroplast genome and contains a co-transcribed region consisting of three highly conserved exons that code the transfer RNA (tRNA) genes for threonine (UGU), leucine (UAA) and phenylalanine (GAA). The region is interspersed by two intergenic spacers and by a group I intron intercalated within the first and second exon of the *trnL_(UAA)* gene. Phylogenetic results obtained with the *trnT-trnF* region (or part of it) should be treated with caution. This is due to the fact that some recent studies (e.g. Koch et al. 2005; Pirie et al. 2007; Schmickl et al. 2009; Vivjerberg and Bachmann 1999) have shown that there are clearly several copies of certain parts

of this region. If this is ignored, it will easily lead to situations where basic requirement of homology of the characters used for phylogenetic analyses is compromised. This might lead to false hypotheses of phylogeny, especially when they are based on the analyses of only this region.

Larger structural changes (>50 bp) rarely occur in the plastome. However, duplications of the *rpl2* or *rpl23* genes (Bowman et al. 1988) or even the duplication of tRNAs (pseudogenes) are occasionally reported. The later are extremely rare in angiosperms and so far they have only been described from Asteraceae (Vijverberg and Bachmann 1999; Witzell 1999), Annonaceae (Pirie et al. 2007), Brassicaceae (Ansell et al. 2007; Koch et al. 2007; Tedder et al. 2010) and Juncaceae (Drábková et al. 2004). In our recent study we reported a tandem repeat comprising of two to four pseudogene copies upstream of the original *trnF* gene in four *Solanum* (Solanaceae) species (Poczaï and Hyvönen 2011a). We have characterized these structural duplications and shown that they consist of several highly structured motifs, which are partial residues, or entire parts of the anticodon, T- and D-domains of the original gene, but all lack the acceptor stems at the 5' or 3'. We were

* Correspondence: peter.poczaï@gmail.com
Plant Biology, Department of Biosciences, University of Helsinki, PO Box 65,
FIN-00014, Helsinki, Finland

further interested to evaluate the possible occurrence of complete or partial *trnF* pseudogenes in Solanaceae. This family contains many economically important plant species, e.g., potato (*Solanum tuberosum* L.), tomato (*Solanum lycopersicum* L.) and paprika (*Capsicum annuum* L.) and is under intensive phylogenetic investigation and the *trnT-F* plastid marker is commonly used in these studies. These sequences together with the results of molecular breeding programs provide large amount of data that is available in GenBank. During data mining we concentrated on a structured dataset generated in previous phylogenetic studies (Fukuda et al. 2001; Garcia and Olmstead 2003; Santiago-Valentin and Olmstead 2003; Bohns 2004; Clarkson et al. 2004; Levin and Miller 2005; Levin et al. 2005; Weese and Bohns 2007; Olmstead et al. 2008) that contained 195 taxa and 390 sequences. This dataset provided the basis for the latest robust phylogenetic hypothesis of the Solanaceae including 89 from the 98 (Olmstead and Bohns 2007) recognized genera. Manual search using the anticodon domain of the original *trnF* gene and automated tRNA recognition by CENSOR (Kohany et al. 2006) indicated the presence of pseudogene repeats in numerous genera of Solanaceae.

We used the core *trnL-F* dataset to map the occurrence of pseudogenic repeats on the phylogenetic tree of Solanaceae. As presented in Figure 1 the distribution of pseudogenic duplications is in congruence with the previously published phylogeny of the Solanaceae (Olmstead et al. 2008), and it is obvious that the first pseudogenic copy evolved only once at the base of a highly supported clade within the subfamily Solanoideae. Among the members of this lineage, referred here as the Pseudosolanoid clade, the anticodon domain of the *trnF* gene exhibits extensive gene duplications with one to seven tandemly repeated copies in close 5'-proximity of the original functional gene (Table 1). The size of each pseudogenic copy ranged between 32 and 73 bp and the anticodon domain was identified as the most conserved element. A common ATT(G)_n motif is of particular interest and its modifications were found to border the 5' of the duplicated regions in the same way as found in Brassicaceae (Ansell et al. 2007; Koch et al. 2005 and 2007; Schmickl et al. 2009; Tedder et al. 2010). Other motifs were partial residues or entire parts of the T- and D-domains. The residues of the 3' and 5' acceptor stems were rarely found among the copies (see Table 1). The D-domain was more conserved than the T-domain among the copies and other internal repeats (AT, AAT, ATT, AATCC) were intercalated within this region for example in genus *Lycianthes* (Dunal.) Hassl. In addition to these newly discovered pseudogenes we were also able to characterize putative promoter motifs showing high similarity to a sigma⁷⁰-type bacterial promoter. These two elements (-35 TTGACA/-10 GAGGAT) are consistently found in the *trnL-F* spacer region of embryophytes, and they are believed to represent the ancient and original

trnF gene promoter (Quandt et al. 2004). Interestingly, pseudogenic repeats were found to be exclusively inserted after such motifs in Solanaceae, contrary to Brassicaceae, where similar pseudogenic repeats were found only between promoter motifs in the *trnL-F* intergenic spacer region (Koch et al. 2005). The later finding lead Koch et al. (2005) to support the conclusion by Kanno and Hirtai (1993) that these elements should be non-functional due to the intercalated position of pseudogenes between promoters. However, this may be challenged by the position of Solanaceae pseudogenes following the -10 and -35 promoters, which are also variable in number and composition.

The occurrence of pseudogenes provides strong evidence of relationships among some groups that had low support values in the previous analyses (e.g. Olmstead et al. 2008). This event robustly separates the (1) Atropina (Hyoscyameae, Lycieae, Jabrosa, Latua, Nolana and Scleraphylax) and (2) Juanulloae clades from the Pseudosolanoid clade composed by (3) Solanaeae, Capsiceae, Physaleae and Datureae and (4) Salpichroina (*Salpichroa* Miers and *Nectouxia* Kunth). In clades (1) and (2) pseudogenes are absent while they appear at the basal node of clade (3) and (4). This lineage where pseudogene copies have been found includes 29 genera; here belongs also the clade of *Solanum* L. and *Capsicum* L. with many economically important plant species. However, sequence information was lacking for the genera *Mellissia* Hook. f. and *Athenaea* Adans. to confirm the presence of *trnF* pseudogenes. This is not surprising as available plant material of these taxa is very restricted. For example *Mellissia* is a genus with a single species, *Mellissia begoniifolia* (Roxb.) Hook. f. which is critically endangered and endemic to the island of Saint Helena. The larger clade of Solanoideae also includes several branches with low support values composed of small genera (*Exodeconus* Raf., *Mandragora* L., *Nicandra* (L.) Gaerten., *Schultesianthus* Hunz., *Solandra* Sw.) in the phylogeny proposed by Olmstead et al. (2008). These lineages are from the early diversification of the Solanoideae with no close relatives and all lack pseudogene repeats that could be informative to trace their ancestry.

The latest large scale phylogenetic analysis of the Solanaceae (Olmstead et al. 2008) established major clades of the family but sampling in some of the lineages can still be improved. Goldberg et al. (2010) analyzed a larger data set but they did not focus on taxonomic relationships but rather on the evolution of self-compatibility. Some studies have attempted to calibrate a molecular clock for various groups within Solanaceae, but all of these used the same (Paape et al. 2008; Poccai and Hyvönen 2011b), or only few fossil records (Dillon et al. 2009; Tu et al. 2010). Fossil record of the Solanaceae has not been reviewed recently. This

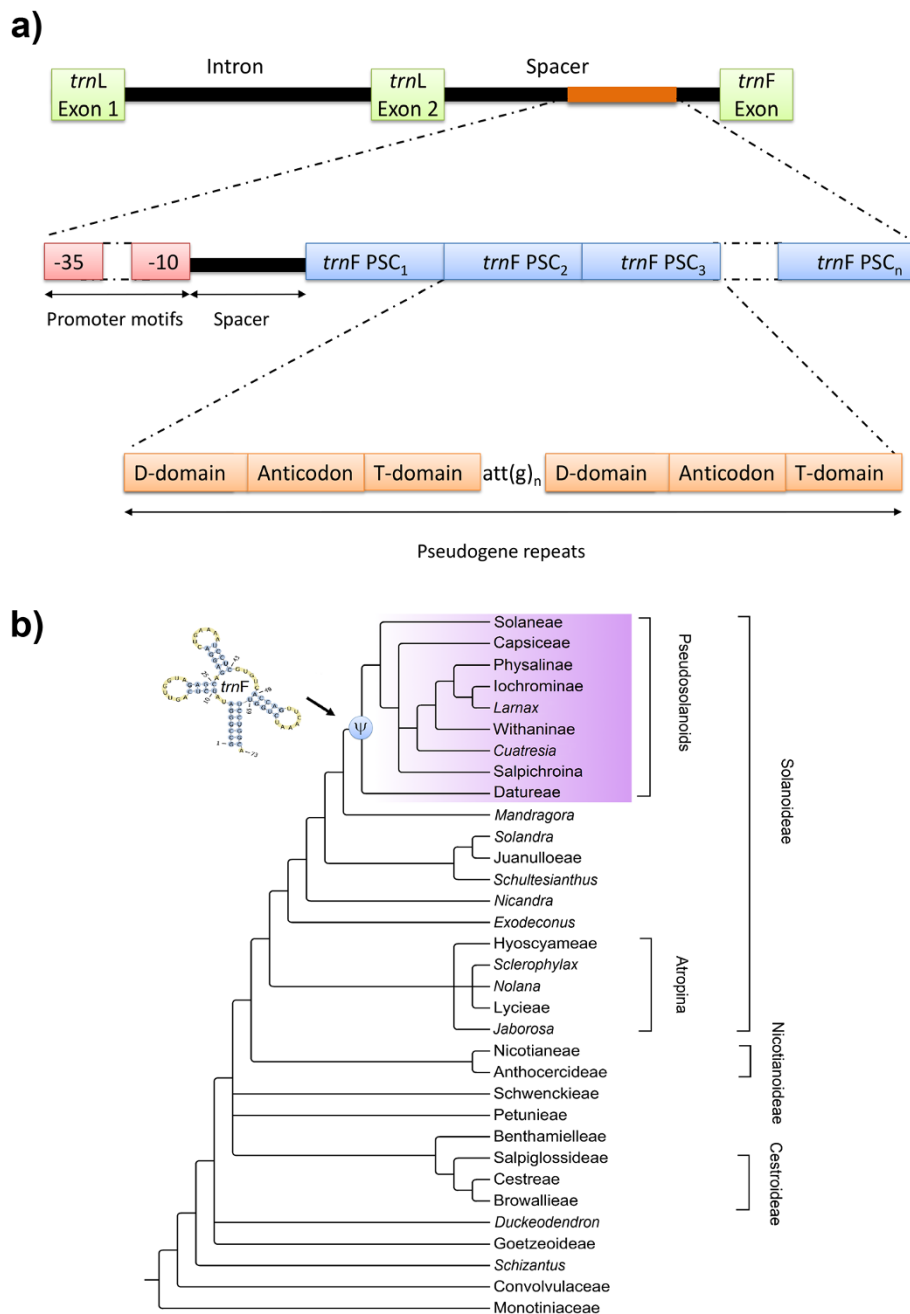


Figure 1 Phylogeny of Solanaceae and the distribution and schematic structure of *trnF* pseudogene copies. **a)** Suprageneric groups recognized are indicated to the right on the tree, while major clades are collapsed at the base node and their names follow Olmstead et al. (2008). The new Pseudosolanoid clade united by the presence of pseudogenetic *trnF* gene duplication is marked with 'ψ' in the Solanoideae subfamily. **b)** The schematic representation of the plastidic *trnL-F* spacer region in Solanaceae and the intercalated pseudogene copies (PSC) in the intergenic spacer region close to 5' of the *trnF* gene. Pseudogene repeats are variable in number and structure and are found after the putative promoter motifs that are also variable among species. The spacer region between the first PSC and promoter motifs consists of intergenic repeats of variable length. Each PSC is separated by a common bordering motif (ATTG) at the 5' end.

urges for the re-assessment of the specimens and could potentially provide more robust calibration points for the family (Särkinen, personal communication). Latest current estimates show the age of the Pseudosolanoids to be

approximately 20 My (Särkinen, personal communication), and thus the origin of the pseudogene duplications of Solanaceae to be approximately of the same Miocene age as in Brassicaceae (16–21 My; Koch et al. 2005).

Table 1 Distribution of *trnF* pseudogenes among Solanaceae and number of multiplicated *trnF* anticodon domains

Taxa	GenBank	Tribe	Copy number
<i>Acnistus arborescens</i>	EU580954	Physaleae	2 ^{a,b}
<i>Aureliana fasciculata</i>	EU580961	Physaleae	2
<i>Brachistus stramonifolius</i>	EU580963	Physaleae	3
<i>Brugmansia aurea</i>	EU580965	Datureae	1 ^c
<i>Brugmansia sanguinea</i>	EU580966	Datureae	1 ^c
<i>Capsicum baccatum</i>	EU580969	Capsiceae	4 ^{a,b}
<i>Capsicum chinense</i>	EU603443	Capsiceae	4 ^d
<i>Capsicum minutiflorum</i>	EU580970	Capsiceae	6 ^d
<i>Capsicum pubescens</i>	AY348982	Capsiceae	6 ^{b,d}
<i>Capsicum rhomboideum</i>	EU580971	Capsiceae	1 ^e
<i>Chamaesaracha coronopus</i>	EU580978	Physaleae	4
<i>Chamaesaracha sordida</i>	EU580979	Physaleae	4
<i>Cuatresia exiguiflora</i>	EU580981	Physaleae	2
<i>Cuatresia riparia</i>	EU580982	Physaleae	2
<i>Datura leichhardtii</i>	EU580983	Datureae	1 ^f
<i>Datura stramonium</i>	EU580984	Datureae	1 ^f
<i>Deprea sylvorum</i>	EU580985	Physaleae	3
<i>Discopodium penninervum</i>	EU580986	Physaleae	4
<i>Dunalia solanacea</i>	EU580988	Physaleae	4
<i>Eriolarynx lorenzii</i>	EU580990	Physaleae	4
<i>lochroma australe</i>	EU580999	Physaleae	4
<i>lochroma cardenasianum</i>	EU581000	Datureae	1 ^f
<i>lochroma fuchsioides</i>	EU581001	Physaleae	2
<i>lochroma umbellatum</i>	EU581002	Physaleae	2
<i>Jaltomata auriculata</i>	EU581006	Solaneae	2 ^f
<i>Jaltomata grandiflora</i>	EU581007	Solaneae	2 ^f
<i>Jalotmata procumbens</i>	AY098695	Solaneae	1 ^{a,b,f}
<i>Jaltomata sinuosa</i>	DQ180418	Solaneae	2 ^f
<i>Larnax subtriflora</i>	EU581009	Physaleae	3
<i>Leucophysalis grandiflora</i>	EU581013	Physaleae	2
<i>Leucophysalis nana</i>	EU581014	Physaleae	2
<i>Lycianthes biflora</i>	EU581015	Capsiceae	2 ^g
<i>Lycianthes ciliolata</i>	EU581016	Capsiceae	4 ^{h,i}
<i>Lycianthes glandulosa</i>	EU581017	Capsiceae	3 ^{g,i}
<i>Lycianthes heteroclita</i>	DQ180414	Capsiceae	2 ^{g,i}
<i>Lycianthes inaequilatera</i>	EU581018	Capsiceae	6 ^{h,i}
<i>Lycianthes multiflora</i>	EU581019	Capsiceae	3 ^{g,i}
<i>Lycianthes peduncularis</i>	EU581020	Capsiceae	4 ^{h,i}
<i>Lycianthes shanesii</i>	EU581021	Capsiceae	1 ^{g,h,i}
<i>Margaranthus solanaceus</i>	EU581025	Physaleae	5 ⁱ
<i>Nectouxia formosa</i>	EU581031	Salpichroina*	1 ^{a,b}
<i>Nothoecstrum latifolium</i>	EU581037	Physaleae	2

Table 1 Distribution of *trnF* pseudogenes among Solanaceae and number of multiplicated *trnF* anticodon domains (Continued)

Taxa	GenBank	Tribe	Copy number
<i>Nothoecstrum longifolium</i>	EU581038	Physaleae	3
<i>Oryctes nevadensis</i>	EU581039	Physaleae	3
<i>Physalis alkekengi</i>	DQ180420	Physaleae	2
<i>Physalis carpenteri</i>	EU581042	Physaleae	2
<i>Physalis heterophylla</i>	EU581043	Physaleae	2
<i>Physalis peruviana</i>	EU581044	Physaleae	4 ^{a,b}
<i>Physalis philadelphica</i>	EU581045	Physaleae	5 ^{a,b}
<i>Quincula lobata</i>	EU581051	Physaleae	1 ^{a,b}
<i>Salpichroa origanifolia</i>	EU581052	Salpichroina*	2 ^{a,b}
<i>Saracha punctata</i>	EU581053	Physaleae	4
<i>Solanum abutiloides</i>	AY266236	Solaneae	1 ^f
<i>Solanum aviculare</i>	HM006836	Solaneae	2 ^{a,b,f}
<i>Solanum betaceum</i>	DQ180426	Solaneae	1 ^f
<i>Solanum dulcamara</i>	HM006840	Solaneae	1 ^f
<i>Solanum herculeum</i>	DQ180466	Solaneae	2 ^f
<i>Solanum lycopersicum</i>	NC007898	Solaneae	1 ^f
<i>Solanum melongena</i>	EU176149	Solaneae	2 ^{h,i}
<i>Solanum pseudocapsicum</i>	DQ180436	Solaneae	1
<i>Solanum torvum</i>	AY266246	Solaneae	4 ⁱ
<i>Solanum trisetum</i>	JN130370	Solaneae	2
<i>Solanum wendlandii</i>	DQ180440	Solaneae	1
<i>Tubocapsicum anomalum</i>	EU581066	Physaleae	7
<i>Vassobia dichotoma</i>	EU581067	Physaleae	4
<i>Witheringia cuneata</i>	EU581070	Physaleae	2
<i>Witheringia macrantha</i>	EU581071	Physaleae	5
<i>Witheringia meiantha</i>	EU581072	Physaleae	4
<i>Witheringia mexicana</i>	EU581073	Physaleae	5
<i>Witheringia solanacea</i>	EU581074	Physaleae	3

Taxonomic classification and a GenBank accession number is provided for each species. *Unranked informal clade name.

^aPartial pseudogenic copy at the 3' end. ^bMissing original *trnF* gene. ^cIntact 5' acceptor stem present. ^dThree copies of -35 promoter (TTGACA) motif. ^eFour copies of -35 promoter (TTGACA) motif. ^fOne copy of -10 promoter (GAGGAT) motif present. ^gPseudogene repeats are separated by long internal repeats after the promoter motifs. ^hOnly one copy of -35 promoter (TTGACA) motif. ⁱ3' acceptor stem present.

Conclusions

Despite of the extensive studies based on sequence level characters the taxonomy of the Solanaceae is not yet completely understood. However, there is ongoing work on different levels by multiple groups to resolve phylogenetic relationships (Fukuda et al. 2001; Garcia and Olmstead 2003; Santiago-Valentin and Olmstead 2003; Bohs 2004; Clarkson et al. 2004; Levin and Miller 2005; Levin et al. 2005; Weese and Bohs 2007; Olmstead et al. 2008). There are a number of questions that should be answered regarding the discovery of *trnF* pseudogenes, for example:

How did the duplications originate? Are the pseudogene copy numbers a useful character for phylogenetic inference? To what extent does the number of pseudogene copies vary within a single species? The evolution and structure of pseudogenic copies should be compared with others reported from different plant families especially from Brassicaceae. The potential of *trnF* pseudogenes as phylogenetic markers need to be investigated further in the future for better understanding of the evolution of Solanaceae. These investigations could answer what are the wider implications of the pseudogene repeats for Solanaceae studies that utilize the *trnL-F* spacer region.

Methods

Solanaceae sequence dataset

For the Solanaceae and several outgroups we used the *trnL-F* spacer data assembled by Olmstead et al. (2008). This dataset contained 195 taxa and 390 sequences generated in previous phylogenetic studies (Fukuda et al. 2001; Garcia and Olmstead 2003; Santiago-Valentin and Olmstead 2003; Bohs 2004; Clarkson et al. 2004; Levin and Miller 2005; Levin et al. 2005; Weese and Bohs 2007; Olmstead et al. 2008) and this was used to align and mask pseudogenic copies. The goal was to map the taxonomic distribution of pseudogenes at family level sampling as many genera as possible. This dataset and representative trees used in our study were previously deposited in TreeBASE (ID S2191). This alignment was also used to demonstrate copy number distribution corresponding to the published phylogenetic hypothesis that was not only based on the *trnL-F* spacer information but relied on sequence data from the *ndhF* region.

Recognition and copy number assessment of the *trnF*_(GAA) pseudogenes

The complete chloroplast genome of *Solanum bulbocastanum* Dunal (DQ347958) was used to select the corresponding loci of the *trnL-trnF* spacer region (bp positions 48,854 to 49,382), to annotate ambiguous sequences regions, and to ensure that our interpretations are based on homologous positions. Putative pseudogene repeats were identified with screening using Repbase (Jurka 2000) with the “mask pseudogenes” and “report simple repeats” options of the online tool CENSOR (Kobany et al. 2006). This was done to identify repetitive elements by comparing our sequences to known eukaryotic repeats and prototypic sequences stored in Repbase utilizing WU-BLAST. A second search was conducted with FastPCR (Kalendar et al. 2009) using the repeat search option of the program. Under “type of repeats” we checked for simple, direct, inverted, direct anti-sense, and direct reverse repeats, respectively. Default values were used under a kMers repeat screening. After each search, repetitive motifs and sequences were

recorded and compared with the results obtained from the Repbase search. After repeats were identified in the *trnL-F* IGS sequences, further structural *trnF*_(GAA) gene elements or residues were annotated manually using the anticodon domain as reference. The annotated sequence alignment is shown in Additional file 1.

Sequence annotation and alignment

Masked pseudogenic copies were further edited using Geneious v.4.8.5 (Biomatters Ltd.). We used the *Nicotiana tabacum* L. complete chloroplast genome (NC001879; bp positions 49,840 to 50,318) for comparisons and to determine the subunits of pseudogenic repeats as this species lacks these gene duplications. Sequence break points were examined manually to determine the cut off points of pseudogenic copies and to identify bordering motifs. Identified copies were aligned with MUSCLE (Edgar 2004) as implemented in Geneious v.4.8.5 using default settings. The sequence alignment in FASTA format is available as Additional file 2.

Additional files

Additional file 1: Annotated sequence alignment of pseudogene repeats found in Solanaceae. Major parts of the *trnF* gene are marked as D- and T-domains and anticodon in the middle together with bordering 5' and 3' acceptor stems. The *trnF* gene of *Nicotiana tabacum* is used as a reference sequence to align different pseudogenes.

Additional file 2: Sequence alignment of pseudogene copies.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PP conceived the study and drafted the early version of the manuscript and performed database search. JH commented on the manuscript, revised the text and structure, and outlined it several times together with PP. Both authors approved the final manuscript.

Acknowledgments

PP gratefully acknowledges support from the Marie Curie Fellowship Grant (PIEF-GA-2011-300186) under the seventh framework program of the European Union. We thank Neil Bell for discussions on the manuscript.

Received: 27 June 2013 Accepted: 11 September 2013

Published: 12 September 2013

References

- Ansell SW, Schneider H, Pedersen N, Grunmann M, Russell SJ, Vogel JC (2007) Recombination diversifies chloroplast *trnF* pseudogenes *Arabidopsis lyrata*. *J Evol Biol* 20:2400–2411
- Bohs L (2004) A chloroplast DNA phylogeny of *Solanum* section *Lasiocarpa*. *Syst Bot* 29:177–187
- Bowman CM, Barker RF, Dyer TA (1988) The location and possible evolutionary significance of small dispersed repeats in wheat ctDNA. *Curr Genet* 10:931–941
- Clarkson JJ, Knapp S, Garcia VF, Olmstead RG, Leitch AR, Chase MW (2004) Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. *Mol Phylogenet Evol* 33:75–90
- Dillon MO, Tu T, Xie L, Quipuscoa Silvestre V, Wen J (2009) Biogeographic diversification in *Nolana* (Solanaceae), a ubiquitous member of the Atacama and Peruvian Deserts along the western coast of South America. *J Syst Evol* 47:457–476

- Drábkova L, Kirschner J, Vlček Č, Pačák V (2004) TrnL-trnF intergenic spacer and trnL intron define major clades within *Luzula* and *Juncus* (Juncaceae): importance of structural mutations. *J Mol Evol* 59:1–10
- Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
- Fukuda T, Yokoyama J, Ohashi H (2001) Phylogeny and biogeography of the genus *Lycium* (Solanaceae): inferences from chloroplast DNA sequences. *Mol Phylogenet Evol* 19:246–258
- García VF, Olmstead RG (2003) Phylogenetics of tribe Anthocercideae (Solanaceae) based on ndhF and trnL/F sequence data. *Syst Bot* 28:609–615
- Goldberg EE, Kohn JR, Lande R, Robertson KA, Smith SA, Igic B (2010) Species selection maintains self-incompatibility. *Science* 330:459–460
- Jurka J (2000) Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet* 9:418–420
- Kalendar R, Lee D, Schulman AH (2009) FastPCR software for PCR primer and probe design and repeat search. *Genes Genomes Genomics* 3:1–14
- Kanno A, Hirtai A (1993) A transcription map of the chloroplast genome from rice (*Oryza sativa*). *Curr Genet* 23:166–174
- Koch MA, Dobeš C, Matschinger M, Bleeker W, Vogel J, Kiefer M, Mitchell-Olds T (2005) Evolution of the trnF(GAA) gene in *Arabidopsis* relatives and the Brassicaceae family: monophyletic origin and subsequent diversification of a plastidic pseudogene. *Mol Biol Evol* 22:1032–1043
- Koch MA, Dobeš C, Kiefer C, Schmickl R, Klimeš L, Lysak MA (2007) Supernetwork identifies multiple events of plastid trnF(GAA) pseudogene evolution in the Brassicaceae. *Mol Biol Evol* 24:63–73
- Kohany O, Gentles AJ, Hankus L, Jurka I (2006) Annotation, submission and screening of repetitive elements in repbase: RepbaseSubmitter and censor. *BMC Bioinf* 7:474
- Levin RA, Miller JS (2005) Relationships within tribe Lycieae (Solanaceae): paraphyly of *Lycium* and multiple origins of gender dimorphism. *Am J Bot* 92:2044–2053
- Levin RA, Watson K, Bohs L (2005) A four-gene study of evolutionary relationships in *Solanum* section *Acanthophora*. *Am J Bot* 92:603–612
- Olmstead RG, Bohs L (2007) A summary of molecular systematic research in Solanaceae: 1982–2006. In: Spooner DM, Bohs L, Giovannoni J, Olmstead RG, Shibata D (eds) *Solanaceae VI: genomics meets biodiversity*. Proceedings of the sixth international Solanaceae conference. *Acta Horticulturae* 745. International Society for Horticultural Science, Leuven, pp 255–268
- Olmstead RG, Bohs L, Migid HA, Santiago-Valentin E, Garcia VF, Collier SM (2008) A molecular phylogeny of the Solanaceae. *Taxon* 57:1159–1181
- Paape T, Igic B, Smith SD, Olmstead R, Bohs L, Kohn JR (2008) A 15-myr-old genetic bottleneck. *Mol Biol Evol* 25:655–663
- Pirie MD, Vargas MPB, Botermans M, Bakker FT, Chatrou LW (2007) Ancient paralogy in the cpDNA trnL-F region in Annonaceae: implications for plant molecular systematics. *Am J Bot* 94:1003–1016
- Poccai P, Hyvönen J (2011a) Identification and characterization of plastid trnF(GAA) pseudogenes in four species of *Solanum* (Solanaceae). *Biotechnol Lett* 33:2317–2323
- Poccai P, Hyvönen J (2011b) Phylogeny of kangaroo apples (*Solanum* subg. *Archaeosolanum*, Solanaceae). *Mol Biol Rep* 38:5243–5259
- Quandt D, Stech M (2004) Molecular evolution of the trnT_{UGU}-trnF_{GAA} region in bryophytes. *Plant Biol* 6:545–554
- Quandt D, Müller K, Stech M, Frahm J-P, Frey W, Hilu KW, Borsch T (2004) Molecular evolution of the chloroplast trnL-F region in land plants. *Monogr Syst Bot Missouri Bot Gard* 98:13–37
- Santiago-Valentin E, Olmstead RG (2003) Phylogenetics of the Antillean Goetzeoideae (Solanaceae) and their relationships within the Solanaceae based on chloroplast and ITS DNA sequence data. *Syst Bot* 28:452–460
- Schmickl R, Keifer C, Dobeš C, Koch MA (2009) Evolution of trnF(GAA) pseudogenes in cruciferous plants. *Plant Syst Evol* 282:229–240
- Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol Biol* 17:1105–1109
- Tedder A, Hoebe PN, Ansell SW, Mable BK (2010) Using chloroplast trnF pseudogenes for phylogeography in *Arabidopsis lyrata*. *Diversity* 2:653–678
- Tu T, Volis S, Dillon MO, Sun H, Wen J (2010) Dispersal of Hyoscyameae and Mandragoreae (Solanaceae) from the New World to Eurasia in the early Miocene and their biogeographic diversification within Eurasia. *Mol Phylog Evol* 57:1226–1237
- Vijverberg K, Bachmann K (1999) Molecular evolution of tandemly repeated trnF (GAA) gene in the chloroplast genomes of *Microseris* (Asteraceae) and the use of structural mutations in phylogenetic analysis. *Mol Biol Evol* 16:1329–1340

- Weese TL, Bohs L (2007) A three-gene phylogeny of the genus *Solanum* (Solanaceae). *Syst Bot* 32:445–463
- Wittzell H (1999) Chloroplast DNA variation and reticulate evolution in sexual and apomictic sections of dandelions. *Mol Ecol* 8:2023–2035
- Zhao T, Wang Z-T, Branford-White CJ, Xu H, Wang C-H (2011) Classification and differentiation of the genus *Peganum* indigenous to China based on chloroplast trnL-F and psbA-trnH sequences and seed coat morphology. *Plant Biol* 13:940–947

doi:10.1186/2193-1801-2-459

Cite this article as: Poccai and Hyvönen: Discovery of novel plastid phenylalanine (trnF) pseudogenes defines a distinctive clade in Solanaceae. *SpringerPlus* 2013 **2**:459.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com