Aquaporins in the wild: natural genetic diversity and selective pressure in the PIP gene family in five Neotropical tree species

Delphine Audigeos[1], Anna Buonamici[2], Laurent Belkadi[1§], Paul Rymer[3], David Boshier[3], Caroline Scotti-Saintagne[1], Giovanni G. Vendramin[2], Ivan Scotti[1]*

[1] INRA UMR 0745 EcoFoG (« Ecologie des forêts de Guyane »), Campus Agronomique, BP709 – 97387 Kourou, French Guiana (France)

[2] CNR Florence

[3] Oxford University

*To whom correspondence should be addressed.

[§]Current address:

Corresponding author:

Ivan Scotti

INRA UMR 0745 EcoFoG (« Ecologie des Forêts de Guyane »), Campus Agronomique, BP709 – 97387
Kourou, French Guiana (France)

Phone: +594 594 329274

Fax: +594 594 324302

E-mail: ivan.scotti@ecofog.gf

Summary

The study of the diversity of functional loci in wild populations is central to the understanding of mechanisms of adaptation and population dynamics in natural species. Spatial and temporal variation in genes underlying the interaction of organisms with their environment is one of the most important evolutionary factors that determine the fate of populations and communities, and its knowledge is fundamental for understanding past and present demographic and selective events and for helping predict the response of ecosystems to environmental change.

Tree species represent major components of intertropical biotas; the outstanding biological diversity of tropical forests, and current threats to their very existence, require efforts to unveil the mechanisms regulating the distribution of tree species across environmental conditions and in space. One approach to tackle the interactions between organisms and their environment is the study of genetic diversity at functional loci, with the goal to infer demographic and selective patterns that may affect, or have affected, genes and populations.

Here we describe genetic diversity in a subgroup of the widespread gene family of Aquaporins, the PIP group, in five Neotropical tree species covering five botanical families. Aquaporins are involved in the regulation of water flow both at cell and at organism level and have been shown to be involved in response to drought; therefore they may play a major role in adaptation in seasonally dry forests throughout the world, and in relation to possible global climate changes. Our results show that there is ample polymorphism at PIP loci, that these genes are under balancing selection at least in some of the studied species, and that levels of selective pressure are unequally distributed across populations, possibly matching environmental differences among study sites.

This is the first study about natural variation and evolutionary forces operating on Aquaporins in any species, and the results we show here suggest that studies of functional genetic loci in natural populations of tropical rainforests is a promising way to access to the mechanisms underlying adaptation to their environment.

Keywords: Carapa; Eperua; Bombacopsis; Virola; selection; demography

Introduction

Within the tropics, water availability is one of the most important environmental factors, with soil fertility, determining tree species richness (ter Steege *et al.*, 2006) and distribution (Swaine, 1996). Although wet tropical regions are characterized by high annual rainfall, seasonality makes it unevenly distributed along the year and tropical humid forest can experience seasonal soil drought (Wright, 1992). The length of the dry season may vary from few days to several months, and during these periods soil water content may become very low. Regions nowadays occupied by luxuriant rainforest have undergone decade- to century-long drier spells in recent historical and geological past (**Hammond 2005**). Tree species' natural range may have changed during those periods, but selective pressure may have acted as well on extant populations. Genetic mechanisms of resistance to drought are therefore expected to have evolved in tropical tree species, and variation for these mechanisms is expected as different species, and populations within species, are adapted to different soil water availability optima. Moreover, as global climate changes, rainfall is expected to decrease and dry season length to increase (Hulme and Viner, 1998). Study of natural evolutionary-genetic diversity is needed, if we want to understand ecological mechanisms underlying the composition of such diverse ecosystems as Neotropical rainforests and to predict species responses to climate change. Research on the mechanisms and the genetic bases of drought stress tolerance have been conducted for decades and several reviews exist (Newton *et al.*, 1991; Ingram and Bartels, 1996), but only few studies exist in natural populations and even fewer in tropical tree species.

Commentaire [IS1]: Ask Damien for a reference

Commentaire [IS2]: **DELPHINE:** is this a worldwide expectation, or is it restricted to some parts of the world?

The molecular basis of drought tolerance is extremely complex and a wide variety of expressional candidate genes has been suggested (**Alexanderson (2005)** in *A. thaliana* and e.g. for trees Chang (1996) and Dubos (2003)). Among protein classes involved in response to drought, and in the regulation of water balance in general, aquaporins are a good candidate for the exploration of genetic diversity in natural populations of non-model species such as Neotropical rainforest trees, because they are ubiquitous, well known and the focus of deep functional studies in plants in general

(**Kaldenhoff et al. 2008**) and in trees in particular (**Cochard et al. 2007**). Indeed, they are presents in both prokaryotes and eukaryotes, have a well conserved structure with six transmembrane helices forming a membrane pore and two loops containing asparagine-proline-alanine (NPA) motifs, and they play a role in water transport as water channels (**Kaldenhoff et al. 2008**). In plants, aquaporins form a large family and have been divided into four subfamilies, based on amino-acid sequence comparison (Johanson *et al.*, 2001). Plasma membrane intrinsic proteins (PIPs) form the most conserved group, and Alexandersson *et al.* (2005) have shown that most PIP transcripts are down-regulated upon gradual drought stress.

In this study we have developed universal primers, from all plant sequences available in public database, to sequence PIP genes in five tropical tree species: *Bombacopsis quinata* (Bombacaceae), *Virola sebifera* (Myristicaceae), *Carapa guianensis* (Meliaceae), and two congeneric species *Eperua falcata* and *Eperua grandiflora* (Caesalpiniaceae). We describe their nucleotide diversity in natural populations and apply standard neutrality tests to detect patterns potentially indicating the action of natural selection. Our results point to the widespread action of balancing selection at these loci in populations of Neotropical rainforest trees.

Materials and methods

Sampling

Commentaire [IS3]: **IVAN, DELPHINE, PAUL, ANNA:** we need to add the origin of the samples (both for cloning and for diversity estimation), and from how many individuals the PCRs for cloning were performed.
Just describe the way the sampling was performed, coordinates will be displayed in a table.

Commentaire [IS4]: See Paul's format for species presentation; shorten it slightly and add ecological characteristics of other species too.

DNA extraction.

For all species except *Bombacopsis quinata*, Total genomic DNA was extracted from cambium or leaf tissues following a CTAB method adapted from Doyle & Doyle 1987 and from Colpaert et al. 2005. Approximately 1 cm² of Silica gel-dried tissue (either from leaf or from cambium) was first frozen in liquid nitrogen and then ground to a powder using a mortar and pestle. 0.25 g of insoluble PVP and 1.4 mL of extraction buffer (100 mM TRIZMA® BASE, 2% CTAB, 1.4 M NaCl, 20 mM EDTA and 0.2% βmercaptoethanol) were added before transferring the mix to a tube. Samples were incubated at 65°C with shaking for 30 minutes, then 600 µL 24:24:1 Dichlromethan/chloroform/isoamylalcohol were added. The tubes were homogenized and centrifuged for 15 minutes at 20°C at 13,500 rpm. Supernatant was transferred to another tube, mixed with 1/10 volume of 3M sodium acetate and 1 mL of cold isopropanol, placed for 1 hour at -20°C and then centrifuged for 15 min at 7500 rpm at 4°C. Pellets were first washed with 70% ethanol then with absolute ethanol and dissolved in 100 µL water. DNA quality was analysed by spectrophotometric analysis at 260 nm and 280 nm or by agarose gel electrophoresis. For *B. Quinata*, Qiagen DNeasy 96 Plant Kit (69181) was used to extract genomic DNA from embryos.

Universal primer design, PCR amplification and DNA sequencing

To isolate PIP genes from several tropical tree species, all nucleic and amino-acids sequences of PIP1 and PIP2 subfamilies (**cit**) available in GenBank were aligned with CLUSTALW. These two subfamilies were chosen because they are found in the plasma membrane, well characterised, and evolutionarily recent and close to each other. Degenerated, universal primers were designed in the most conserved regions (Figure 1, table 1.a). To isolate PCR fragments corresponding to the selected loci, two alternative strategies were used. (a) For *Eperua falcata*, RNA was isolated using the protocol described in Kiefer et al. (2000) and cDNA was obtained using Lambda-ZAP-cDNA Synthesis Kit (Stratagene). PIP genes were amplified from cDNA using the degenerated primers. PCR was carried out in a 25 µL-volume containing 5µl cDNA, 1x *Taq* buffer, 1.6 mM $MgCl_2$, 0.24 mM of each dNTP, 1 U Taq polymerase (all products from Invitrogene) and 0.6 µM of each primer. An initial denaturation at 94°C for 5 min was followed by 35 cycles of (45 s at 94°C, 45 s at 64°C and 45 s at 72°C) and a final extension of 5 min at 72°C after the cycles. To isolate the various PIP sequence obtained, PCR products were cloned into pGEM®-T vector (Promega) and PCRs were performed on 48 colonies with the same protocol as described above, but with a lower annealing temperature (60°C). PCR products were cleaned-up with EXOSAP-IT (USB Corporation). Sequencing reactions was performed with

Commentaire [IS5]: Move to / merge with introduction

BigDye® Terminator v3.1 cycle sequencing kit (Applied Biosystems) in a total volume of 10 µl containing 1.5 µl of Big Dye, 1.5 µl of Buffer, 2 µl of 2 µM primer, 4 µl of cleaned-up PCR product and 1µl of milli-Q water. All fragments were sequenced in both directions. Sequencing reactions were then purified on Sephadex columns (Millipore) and sequence data were obtained on an ABI 3130xl capillary sequencer (Applied Biosystems). (b) for the following species: *Cedrela odorata, Carapa guianensis, Bombacopsis quinata, Schizolobium parahyba, Virola sebifera, Ceiba pentandra* the following protocol was used: genomic DNA was extracted with the Invisorb Spin Plant kit (Invitek); universal primers (table 1.a) were used for PCR in a total volume of 12.5 µL, containing 10 ng of genomic DNA, 1×PCR buffer (Promega, it has 1,5 mM MgCl2 final concentration), 0,2 µM each primer, 0,2 mM each dNTP, 1 U Go Taq polymerase (Promega), 0.8% BSA. The reaction was performed with the following thermal profile: 4 min at 94°C; 35 cycles of (30 s at 94°C, 30 s at 50°C, 40 s at 72°C); and a final extension of 7 min at 72°C. Following the PCR, the amplified products were cloned into TOPO vector (Invitrogen) and then sequenced from both ends using a MegaBace 1000 capillary sequencer (Amersham).

Commentaire [IS6]:
ANNA – check this: is this temperature really so low?

Commentaire [IS7]:
ANNA – sequencing protocol? Is it from colony PCR? Or directly on the plasmid? How was the reaction performed?

Sequence analysis and specific primer design

Base calling and contig assembly were done using Phred and Phrap programs (Sequencing Analysis v5.2, Applied Biosystems; CodonCode aligner v2.0.1, Codoncode Corporation). The consensus sequence of each contig was submitted to TAIR Blast (http://www.arabidopsis.org/Blast/) in order to confirm their identity (table 1.a). For each species, contigs were named according to their closest match in TAIR database (either PIP1 or PIP2) followed by contig number. Sequence information was used for the development of gene-specific and species-specific primer pairs for the amplification of gene portions of genomic DNA (Table 1.b., Figure 2). At this step, 48 double-stranded sequences were used for *E. falcata*, 44 for *C. odorata*, 46 for *C. guianensis* and *B. quinata*, 34 for *C. pentandra*, 40 for *S. parahyba*, and 43 for *V. sebifera*.

Commentaire [IS8]:
DELPHINE – the consensus sequences must be submitted to GenBank / EMBL prior to publication, and the reference number must appear in a table.

The identity of the isolated sequences was checked by BLAST alignments with publicly available databases; the subdivision of genomic fragments in exons and introns (Figure 2) was obtained by alignment of the genomic fragments with publicly available mRNA sequences. Finally, homology between sequences and small RNA families was checked by matching the former to the latter on the Sanger Centre's webpage (http://microrna.sanger.ac.uk/sequences/index.shtml).

Species- and locus-specific primers (table 1.b) were designed in the outermost regions that granted sufficient divergence among contigs within each species, in order to obtain the largest possible fragments with the best specificity. Primer names are derived from clone name followed by a sign corresponding to the direction of the synthesis ("+" for "forward", "-" for "reverse") and a number corresponding to the position of the primer's 5' end relative to the clone's sequence.

Specific PCR amplification

Specific primers were applied to the amplification of specific genomic DNA; primers drawn from *E. falcata* were also tested in the congeneric *Eperua grandiflora*. Genomic DNA of *E.falcata*, *E.grandiflora* and *C. guianensis* was amplified with the following PCR conditions for all primer pairs: PCRs were carried out in a 15 µL-volume containing 15 ng of DNA, 1x *Taq* buffer, 3 mM $MgCl_2$, 0.25 mM of each dNTP, 0.6 U Taq polymerase (all products from New England Biolabs) and 0.5 M of each primer. An initial denaturation at 94°C for 4 min was followed by 35 cycles of (45 s at 94°C, 30 s at the annealing temperature shown in Table 1.a and 1 min 30 s at 72°C) and a 10 min final extension at 72°C after the cycles. For *C. odorata*, *C. pentandra*, *S. parahyba* and *V. sebifera*, PCR reactions were performed in a 12.5 µL reaction volume containing 10 ng of DNA, 1×PCR buffer (Promega), 0.2 µM each primer, 0.2 mM each dNTP, 2.5 mM $MgCl_2$, 1 U Go Taq polymerase (Promega) and 0.8% BSA, using the following thermal profile: initial denaturation at 95°C for 5 min, 30 cycles of (30 s at 95°C, 30 s at 59°C, 30 s at 72°C); final extension at 72°C for 7'. For *B. quinata*, PCRs were performed in a total volume of 10uL reaction containing 1µL 2.5mM BSA, 0.6uL 2.5mM dNTPs, 0.4 µL 50mM $MgCl_2$, 0.2µL 10mM fwd and rev primers, 0.1µL 5U Taq (Yorkshire Biosciences A2002 YB-TAQ DNA

Commentaire [IS9]: **ANNA:** these three species disappear from subsequent analyses, and we should either explain why or remove them from the paper (see also comment IS16).

Commentaire [IS10]: **ANNA:** is this the only temperature you have tested? Please add a sentence or two in the results, explaining how annealing temperatures were optimized (see also first comment in the Results section). However, for the sake of uniformity, this has to be replaced with: " Ta (table 1.b)"

Polymerase). The following thermal profile was applied: 3 min at 94 °C, 30 cycles of (30 s at 94 °C, 30 s at T_a (table 1.b), 1 min at 72 °C), final extension at 72 °C for 10 min.

Commentaire [IS11]:
PAUL: could you please convert these amounts into concentrations, in order to align your protocol to the others?

DNA polymorphism, population structure and demographic processes:

Since the DNA samples were diploid, the identification of haplotypes (*i.e.* sequence variants) was ambiguous where more than one SNP was present and heterozygote individuals were observed. Diploid sequences were treated using Haplotyper© (Harvard University), a software for haplotype inference using the Bayesian algorithm, to produce two haploid sequences per individual. Insertions-deletions ("indels") were coded like SNPs: each gap, irrespective of its length, was replaced by a nucleotide producing a SNP, in order to treat indel in subsequent analyses.

To test population structure, we computed population comparisons, based on pairwise difference (F_{ST}) between geographical populations, then AMOVA tests were performed using ARLEQUIN v3.01 (Excoffier, http://cmpg.unibe.ch/software/arlequin3/, 2006). For some species, two subgroups of samples were detected, that represented genetically homogeneous populations. Since neutrality tests can only be performed on undifferentiated populations, they were performed either at the species- or at the population level according to population differentiation results. Subgroups are displayed as "pop 1" and "pop 2" in Tables 1 and 2.

Commentaire [IS12]:
IVAN, DELPHINE, PAUL: See first comment in M&M – we need to define the sampling sites where necessary

Commentaire [IS13]:
DELPHINE: Check authorship

Commentaire [IS14]:
DELPHINE: F_{ST} and AMOVA results may be displayed in a table that we will add as supplementary material (not shown in the paper, but published on the journal's website)

Analyses of sequence data were performed using DnaSP v. 4.10.9 (Rozas et al., 2003). Nucleotide diversity was estimated by Watterson's θ_w (Watterson 1975) and π, the average number of pairwise nucleotide differences among sequences in a sample (Nei, 1987).

A number of statistical analyses were conducted to identify departures from the standard neutral model of evolution, like Tajima's *D* (1989), Fu and Li's tests (1993) and Fu's F_s (1997). Tajima's *D*-statistic was computed for each locus and reflects the difference between π and θ_W. Fu and Li's tests (*D* and *F*) were computed without outgroup, and consider the distribution of the mutations in the genealogy, comparing "old" and "new" mutations. The F_s test statistic for neutrality (Fu 1997), based on the haplotype (gene) frequency distribution was also calculated. Recombination rate of gene fragments was estimated by LDhat v2.1 (McVean and Auton, 2007, www.stats.ox.ac.uk/~mcvean/LDhat.html) and coalescent simulations with DnaSP were performed with and without recombination, because the accumulation of historical recombination events influences patterns of sequence diversity even on very short genetic distances.

More detailed analyses were performed within each sequence by a sliding-window method; test statistics were recomputed on windows of size = ... base pairs.

Commentaire [IS15]: **DELPHINE:** describe this method more precisely

Haplotype networks were realised, first by generating inter-haplotypic distance matrices with ARLEQUIN (Excoffier, 2006, http://cmpg.unibe.ch/software/arlequin3/), then by computing a minimum spanning network (embedding all minimum spanning trees) with MINSPNET program (Excoffier and Smouse, 1994).

Commentaire [IS16]: **DELPHINE:** Check authorship

The same samples analysed for sequence diversity were also genotyped at eight chloroplast loci using universal primer pairs (ccmp1, ccmp2, ccmp3, ccmp4, ccmp5, ccmp6, ccmp7, ccmp10; CIT); the patterns of diversity obtained on these loci were analysed with the “sign test” method included in the BOTTLENECK software package (Piry et al. 1999) in order to test for (recent) demographic events in the populations, that may be detected by tests of selection on sequences and thus confound the results. The results obtained on chloroplast microsatellites were used as the neutral reference, undergoing demographic transitions only (if any), based on which the departures from neutral equilibrium, obtained on PIP sequences, were ascribed to demography or to selection. The Stepwise Mutation Model and 1000 replications were used for these tests.

Commentaire [IS17]: **ANNA**: please provide appropriate reference

Results

Amplifications with universal primers have allowed the cloning of 1-5 different genes per species and the sequence information thus obtained was used for the development of gene-specific and species-specific primer pairs for the amplification of seven gene portions (Table 1). We sequenced a total of

Commentaire [IS18]: **DELPHINE, ANNA:** more details should be given on the results of isolation of gene fragments, as this is one of the most original pieces of this paper: how many contigs per species? Which is their closest homolog in TAIR? A table resuming all these pieces of information should be added.
Moreover, PCR products with universal primers were amplified from a larger set of species than shown in table 1.b, so where have all other species gone? We must explain this explicitly. Same for the contigs that, for one reason or another, did not make it to polymorphism analysis. Finally, the result of the transfer of PCR primers from *E. falcata* to *E. grandiflora* should be described here.

Commentaire [IS19]: **DELPHINE:** this seems to correspond to the sequences that have been characterised *at the population level*, as opposed to the total length of the identified contigs; can you explain and/or add total contig length?

4036 bp both in coding and non-coding regions (Figure 2). Polymorphism distribution in each gene fragment and each species is shown in table 2. Polymorphism was identified in all species, although its amount varies according to the gene and the species. A total of 79 SNPs (including indels) was found, distributed across coding and non-coding regions and between synonymous and non-synonymous sites.

The geographical distribution of haplotypes was tested for each species and each gene in order to identify homogeneous populations. For two species (*C. guianensis* and *E. falcata*), a subdivision was detected (Supplementary Figure 1 and Supplementary Table 1) and taken into account for all subsequent analyses. This allowed us to perform tests on the demographic/selective significance of the observed polymorphism patterns, taking into account recombination rates (Table 3). For five genes/populations, the detection of recombination events gave significant results: CguPIP1.3 pop 2, VsePIP2.5, BquPIP2.4 and EfaPIP1.1 pop 1 and EgrPIP2.1. Recombination rate was null for CguPIP1.3 pop 2 and equal to 1 for EgrPIP2.1.

Commentaire [IS20]:
DELPHINE: this figure needs some refinement but is generally very good

Commentaire [IS21]:
DELPHINE: let us only show data with recombination (see also comment to table 3)

Commentaire [IS22]:
DELPHINE: what does this mean?

Three gene fragments/populations out of eight did not reveal any departure from neutrality (EfaPIP1.1 pop2, EfaPIP1.2, EfaPIP2.1).

For CguPIP1.3 pop 2, all neutrality tests were positive and significantly different from zero, due to excess of variants at intermediate frequencies (Tajima's *D*), deficiency of recent mutations (Fu and Li's *D** test) and a lower number of haplotypes than expected (Fu' F_s). This pattern of polymorphism is consistent both with a contraction of effective population size and with balanced selection. Population 2 of *C. guianensis* contains only two haplotypes, separated by eight mutations (supplementary Figure 2) but none of them leads to non-synonymous changes.

For EgrPIP2.1, only Fu and Li's *D**-test were positive and significant, suggesting a deficit of recent mutations and indicating that mutations have occurred in the older part of the genealogy, so that the selective or demographic event responsible for this patterns is relatively old.

Commentaire [IS23]:
DELPHINE: should this be interpreted in the same way as CguPIP1.3?

For the other genes (BquPIP2.4, VsePIP2.5 and EfaPIP1.1 pop 1) test significance changed when recombination was taken into account.

For *B. quinata* gene BquPIP2.4, Fu and Li's *D**-test was positive and significant with or without recombination but Fu's F_s was positive and became significant only when recombination rate was taken into account. These results are associated with a deficit of recent mutations and a lower number of haplotypes than expected, so the data can be interpreted the same way as for *C. guianensis*.

For VsePIP2.5, Fu's *F*s-test was negative only when recombination was not taken into account, indicating that the departure from neutrality is only apparent.

For EfaPIP1.1 pop 1, Fu's F_S became significant when recombination was taken into account, due to population growth or hitchhiking (Fu, 1997). In this case, F_S is negative and significant, but higher than the expected values, so it has to be interpreted as a *positive and significant* test. Indeed, when considering only mutations and genetic drift but not recombination, simulated values of F_S are centred on zero, but when recombination rate is taken into account, the expected values become negative. So the interpretation of this test corresponds to a reduction in population size or balanced selection. Interestingly, a sliding-window analysis of the neutrality tests (Figure 3) shows that most windows containing informative polymorphisms share positive test statistic values, except for one window overlapping the intron (which starts at base 180 of this PCR fragment) and centred on bases 286-288. When matched against small RNA databases, the intron fragment contained in EfPIP1.1 sequences turns out to be similar to immature *Arabidopsis thaliana* miRNA ath-MIR863 and to *Oryza sativa* miRNA osa-MIR420 (with Evalues of 0.048 and 0.064, respectively).

Commentaire [IS24]: DELPHINE: let us discuss this point

Results on demographic transitions, as detected on chloroplast SSRs, are shown in table 4. Globally, no species or population displayed signatures of population *contraction*, and most species and populations displayed signatures of population *expansion* instead. *C. guianensis* pop 2, *V. sebifera* and *B. quinata* did not display departures from mutation / drift equilibrium, but the tests could only be performed on one locus due to extensive lack of polymorphism. For *E. falcata* Pop 1 and for *E. grandiflora*, signatures of population expansion are found at SSR loci concurrently with signatures of population contraction and/or balancing selection at PIP loci. *C. guianensis* Pop 1 shows signatures of population expansion but was monomorphic for PIP sequences. *E. falcata* Pop 2 also has a signature for population expansion, but no neutrality test was significant.

Discussion

We provide here a method for the rapid isolation of gene family members from non-model species, based on sequence information drawn from public databases. In particular, we have developed a universal primer pair which makes it possible to explore natural gene diversity in the PIP sub-family of aquaporins. This study reports nucleotide diversity for six PIP gene fragments in five tropical tree species; it demonstrates that the sequences thus obtained are useful for the detection of sequence polymorphism and can be used for testing departures from the neutral theory of molecular evolution (Kimura 1983). In one case (PIP2.1) the same primer pair was used to amplify two congeneric species (Table 1.b), thus providing a direct comparison between orthologs. For these two genes (EfaPIP2.1 and EgrPIP2.1, table 3) the amount of polymorphism in *E. falcata* is lower than in *E. grandiflora*, and departure from neutrality was detected in the latter species but not in the former. Analyses which require an outgroup are possible like HKA-test (**Hudson *et al*., 1987**) and MK-test (**McDonald and Kreitman, 1991**).

Commentaire [IS25]: **DELPHINE:** in the Results section (see first comment in the Results section) we should have a look at what the non-synonymous mutations do, to make sure that we are not comparing paralogs or, worse, a pseudogene in *E. grandiflora*

Commentaire [IS26]: **DELPHINE, CAROLINE, IVAN:** we should do these tests rather than just say that they are possible.

A number of neutrality tests were conducted to identify genes departing from neutral patterns. Several genes gave significant results. The general trend, for genes and populations for which significant departures from neutrality was observed, is towards reduction of population size and/or balancing selection; however, sliding window analyses show that this trend does not apply to the entire sequence and that different regions of the genes may be subject to divergent processes (Figure 3). Interpretation of these results is ambiguous because significant values can be caused both by selective pressure and by demographic events. However, the results of tests performed on chloroplast SSR loci indicate generally the presence of past population expansion events, thus tending to exclude that the departures from equilibrium observed at PIP loci are due to demographic events, that should be, in the case of PIP loci, contractions instead of expansions. As demography affects equally all portions of the genome, it is unlikely that events of opposing directions are detected on different loci, and therefore we can tentatively conclude that the departures from equilibrium observed at PIP loci are due to balancing selection. A *caveat* should however be kept in mind, as SSRs in general have different mutation rates and mechanisms than coding sequences, and therefore the comparison of results on these two types of data may be misleading. In particular, SSRs tend to mutate more quickly, and the demographic events that we observe at SSR may be more recent than the result of forces acting on sequence diversity. However, we can argue that a recent demographic event, detected on SSR, should be detectable on sequences as well. Tests going opposite ways seem to strongly support selection on the expressed loci. A secondary concern is that effective population size is not the same for chloroplast and nuclear loci, making the former more sensitive to demographic change than the latter. However, again tests going opposite ways should be taken as hints of the action of selection, while cases where demographic signatures are detected at chloroplast markers but not at PIP loci should be conservatively attributed to differences in sensitivity.

Among the gene fragments that, according to neutrality tests, experience selection, only EgrPIP2.1 contains non-synonymous mutations, and therefore, for the remaining four loci, selection may not be acting directly on protein sequence and structure, although it may affect other properties of the transcribed sequence, such as codon composition or intron functions (**ref.**); because PIP sequences are conserved among and within species (Johanson, 2001), selection could rather act on regulatory regions. Alternatively, these polymorphism patterns may reflect selection on neighbouring sites, although the estimated population recombination parameters hint that recombination would quickly break down associations between selected sites and associated neutral sites, so that signatures of

Commentaire [IS27]: **DELPHINE:** we should put these estimates somewhere, perhaps in table 3?

Commentaire [IS28]: **IVAN:** have a closer look at this selection may not extend beyond few hundred base pairs from the site under selection (**ref.**). For genes that did not display any departure from neutrality, full-length sequencing, including the promoter region, is advisable. In general, evolutionary patterns may diverge among different parts of a gene (**Wu et al. 1999**) and **Olsen *et al.* (2002)** found evidence of balanced selection in promoter region in the *TFL1* gene of *Arabidopsis thaliana*.

For *Carapa guianensis* population 2, all neutrality tests on PIP loci gave strong significant results, while no chloroplast SSR marker did. This polymorphism pattern and in particular result of Fu and Li's *D**-test (**Fu and Li, 1993**), is consistent with the maintenance of a balanced polymorphism by a long-term action of balancing selection. Moreover, the two haplotypes observed in population 2 show a geographical pattern (Supplementary figure 1). It is interesting to note that the eastern samples have been collected from a hybridisation zone between *C. guianensis* and the congeneric *C. procera*. The results of Bayesian assignation methods applied to independent loci (SSRs) show, however, that these trees belong to *C. guianensis* populations (Duminil et al. 2006), perhaps pointing to a stable and selectively advantageous introgression of *C. procera* genes into a *C. guianensis* genetic background. A geographical pattern was also detected for one locus for *Eperua falcata* (EfaPIP1.1; Supplementary figure 1) and for the *E. grandiflora* locus EgrPIP2.1 (Supplementary figure 2). The population named "Route de l'est" was removed from analyses because it contained few samples and was separated from the other haplotypes by 8 mutational steps. All subgroups are subdivided by a Northeast-Southwest border running approximately perpendicular to the Atlantic coast in French Guiana, or alternatively show a highly differentiated population at the East of the sampled area. More studies and an extended sampling may reveal other species following the same pattern, perhaps caused by a selective gradient or by common demographic history, shaped by long-term evolution of forest communities. However, direct comparison between the two *Eperua* species for the same locus (PIP2.1) does not reveal the same pattern (EfaPIP2.1 lacks any geographical structure). It will be interesting to extend this comparison to other pairs of sister species. Among the cases described in this study, however, balancing selection seems to be the most common trend, but

Commentaire [IS29]: **DELPHINE:** this statement should be supported by sequencing some samples of *C.procera*.

at the present time it is hard to outline a reason as to why this is the case. One possibility is that the sampled populations may be further structured along ecological gradients at a smaller geographic scale, and that different environments favour alternative alleles, thus maintaining genetic diversity within populations. Alternatively, large environmental variations, both seasonal or over longer cycles, may prevent selection from fixing a particular variant. Thus, variation of selective regime both in space and in time may underlie these patterns.

For the other genes, interpretation of the neutrality tests is not obvious, because we cannot avoid recombination event, especially to study natural populations. Indeed, most of neutrality tests are conservative because they assume no intralocus recombination (**Tajima, 1989; Fu and Li, 1993; Fu, 1997**). Fu (**1996**) note that recombination events do not change the expectations of θ_π and θ_w estimators but reduce their variances and increase the number of alleles in a sample. The larger the number of alleles in a sample is, the less likely the neutral model will be rejected by tests based on these estimators (like Tajima'*D* and Fu and Li *D** and *F**). Therefore, test *F*s may be sensitive to recombination and application of this test may be inappropriate.

In order to test the presence of selection, further studies could be established to test association between haplotype and adaptive traits related to water stress. Moreover, investigations for selection in other candidate genes may be conducted. Indeed, a large number of genomic or proteomic studies have identify candidate gene for water stress tolerance. For example dehydrin genes are up-regulated by drought stress [*BjDHN2* and *BjDHN3* in *Brassica juncea* (Xu *et al.*, 2008); *PgDhn1* in *Picea glauca* (Richard *et al*, 2000)] and alcohol dehydrogenase (*Adh*) transcripts are induced by anoxia and hypoxia (Sachs *et al*., 1980; Gregerson et al. 1993). Contrasting the results obtained on these genes with analyses performed on neutral loci to detect demographic events will also be fundamental to break apart the two intertwined interpretations (demographic versus adaptive) of tests for departure from neutrality.

The fragments of PIP (aquaporin) genes analysed here have revealed large variability and potentially strong signatures of past population-genetic events, in some cases with patterns that vary along the gene. An extended characterisation of these loci is likely to reveal more details on the processes shaping their diversity and to provide information on the link between genetic diversity and ecological conditions. Systematic characterisation of (fragments of) candidate genes, facilitated by the gene-cloning method developed here, may lead us to a more complete picture of the way genotypes interact with environment in tropical rainforest ecosystems. We are convinced that this is a necessary step towards the consolidation of ecological genetics in tropical ecology.

Acknowledgements

This work was funded by the EU-funded INCO "SEEDSOURCE" program and by the French Ministry of Ecology and Sustainable Development "ECOFOR – ECOSYSTEMES TROPICAUX" program. The authors wish to thank Pauline Garnier-Géré for methodological discussions and help, and Saint-Omer Cazal and Jean Weigel for sampling, and ... for critically reading the manuscript.

References

Commentaire [IS30]: The list is still incomplete; if you cannot find a reference you want to have a look at, just ask for it.

Colpaert N, Cavers S, Bandou E, Caron H, Gheysen G, Lowe AJ. 2005. Sampling Tissue for DNA Analysis of Trees: Trunk Cambium as an Alternative to Canopy Leaves. *Silvae genetica* **54**(6): 265-269.

Doyle JJ, Doyle JL. 1987. A Rapid DNA Isolation Procedure from Small Quantities of Fresh Leaf Tissues. *Phytochemistry Bulletin.* **19**: 11-15.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (Version 3.0): An Integrated Software Package for Population Genetics Data Analysis. *Evolutionary Bioinformatics Online* **1**: 47-50.

Excoffier L, Smouse PE. 1994. Using Allele Frequencies and Geographic Subdivision to Reconstruct Gene Trees within a Species - Molecular Variance Parsimony. *Genetics* **136**(1): 343-359.

Fu YX. 1996. New Statistical Tests of Neutrality for DNA Samples from a Population. *Genetics* **143**(1): 557-570.

Fu YX. 1997. Statistical Tests of Neutrality of Mutations against Population Growth, Hitchhiking and Background Selection. *Genetics* **147**(2): 915-925.

Fu YX, Li WH. 1993. Statistical Tests of Neutrality of Mutations. *Genetics* **133**(3): 693-709.

Gregerson RG, Cameron L, McLean M, Dennis P, Strommer J. 1993. Structure, Expression, Chromosomal Location and Product of the Gene Encoding Adh2 in Petunia. *Genetics* **133**(4): 999-1007.

Hudson RR, Kreitman M, Aguade M. 1987. A Test of Neutral Molecular Evolution Based on Nucleotide Data. *Genetics* **116**(1): 153-159.

Hulme M, Viner D. 1998. A Climate Change Scenario for the Tropics. *Climatic Change* **39**(2-3): 145-176.

Ingram J, Bartels D. 1996. The Molecular Basis of Dehydration Tolerance in Plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **47**: 377-403.

Johanson U, Karlsson M, Johansson I, Gustavsson S, Sjovall S, Fraysse L, Weig AR, Kjellbom P. 2001. The Complete Set of Genes Encoding Major Intrinsic Proteins in Arabidopsis Provides a Framework for a New Nomenclature for Major Intrinsic Proteins in Plants. *Plant physiology* **126**(4): 1358-1369.

Kiefer E, Heller W, Ernst D. 2000. A Simple and Efficient Protocol for Isolation of Functional Rna from Plant Tissues Rich in Secondary Metabolites. *Plant Molecular Biology Reporter* **18**(1): 33-39.

McDonald JH, Kreitman M. 1991. Adaptive Protein Evolution at the *Adh* Locus in *Drosophila*. *Nature* **351**: 652-654.

Nei M. 1987. *Molecular Evolutionary Genetics* New York: Columbia University Press.

Newton RJ, Funkhouser EA, Fong F, Tauer CG. 1991. Molecular and Physiological Genetics of Drought Tolerance in Forest Species. *Forest Ecology and Management* **43**(3-4): 225-250.

Niu. 2006. Bayesian Haplotype Inference for Multiple Linked Single-Nucleotide Polymorphisms (Vol 70, Pg 157, 2002). *American Journal of Human Genetics* **78**(1): 174-174.

Niu TH, Qin ZHS, Xu XP, Liu JS. 2002. Bayesian Haplotype Inference for Multiple Linked Single-Nucleotide Polymorphisms. *American Journal of Human Genetics* **70**(1): 157-169.

Olsen KM, Womack A, Garrett AR, Suddith JI, Purugganan MD. 2002. Contrasting Evolutionary Forces in the Arabidopsis Thaliana Floral Developmental Pathway. *Genetics* **160**(4): 1641-1650.

Piry S, Luikart G, Cornuet J-M 1999. BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of heredity* **90**(4): 502-503.

Richard S, Morency M-J, Drevet C, Jouanin L, Séguin A. 2000. Isolation and Characterization of a Dehydrin Gene from White Spruce Induced Upon Wounding, Drought and Cold Stresses. *Plant molecular biology* **43**: 1-10.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. Dnasp, DNA Polymorphism Analyses by the Coalescent and Other Methods. *Bioinformatics* **19**(18): 2496-2497.

Sachs MM, Freeling M, Okimoto R. 1980. The Anaerobic Proteins of Maize. *Cell* **20**(3): 761-767.

Swaine MD. 1996. Rainfall and Soil Fertility as Factors Limiting Forest Species Distributions in Ghana. *The Journal of Ecology* **84**(3): 419-428.

Tajima F. 1989. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* **123**: 585-595.

ter Steege H, Pitman NCA, Phillips OL, Chave J, Sabatier D, Duque A, Molino JF, Prevost MF, Spichiger R, Castellanos H, von Hildebrand P, Vasquez R. 2006. Continental-Scale Patterns of Canopy Tree Composition and Function across Amazonia. *Nature* **443**(7110): 444-447.

Watterson GA. 1975. On the Number of Segregating Sites in Genetical Models without Recombination. *Theoretical population biology* **7**(2): 256-276.

Wright SJ. 1992. Seasonal Drought, Soil Fertility and the Species Density of Tropical Forest Plant-Communities. *Trends in Ecology & Evolution* **7**(8): 260-263.

Xu J, Zhang YX, Guan ZQ, Wei W, Han L, Chai TY. 2008. Expression and Function of Two Dehydrins under Environmental Stresses in Brassica Juncea L. *Molecular breeding* **21**(4): 431-438.

Tables

Table 1. Description of the PIPs fragments amplification conditions. Primer names, primer sequences, annealing temperature, size of the sequenced fragment, Accession number of the closest *Arabidopsis thaliana* Blast match.

Commentaire [IS31]:
DELPHINE, ANNA: This table appears here as a figure, and needs to be back-converted into a table. Primer names should appear as complete names, including primer positions and not only clone name. Moreover: in Paul's table, BquPIP2.1 and BquPIP2.2 appear, whereas in Delphine's and Anna's table BquPIP2.4 appear. Please clarify and merge all information into a single table.
Some more minor modifications are also needed.

	Species	*Primer Name*	*Primer sequences (5' → 3')*	*T_m (°C)*	*Size (bp)*	*TAIR Blast*
a.	All	PIP2H2.2 PIP2H6.1	F: CTYGTYTACTGCACHGCY R: CCVACCCARAADATCCAN	64	850	PIPs mix
b.	*Carapa guianensis*	CguPIP1.3	F: CGGCATTTCAGGTCATCTC R: CCAACCCAGAAAATCCAGTG	54	760	AT1G01620
	Bombacopsis quinata	BquPIP2.4	F: GCCGGTATCTCTGGTGAGTG R: CCACGCCTTCTCTTTGTTGT	64	650	AT3G61430
	Virola sebifera	VsePIP2.5	F: CGCGTATCTCTCTCTTCAACG R: CACACGCACACACACAATG	59	750	AT3G54820
	Eperua falcata	EfaPIP1.1	F: CCCAGCAGTGACCTTCG R: AACCAAGAACACAGCGAACC	64 → 57 [1]	550	AT3G61430
		EfaPIP1.2	F: CAACCCGGCTGTGACC R: GCCAAATGGACCAAGAACAC	64 → 57 [1]	550	AT2G45960
		EfaPIP2.1 [2]	F: GCACATAAATCCGGCAGTG R: CCGACCCAGAAGATCCAC	64 → 57 [1]	650	AT3G53420

[1] PCR Touchdown: the first seven cycles with one degree decrease each cycle, from 64°C to 57°C. Other cycles at 57°C

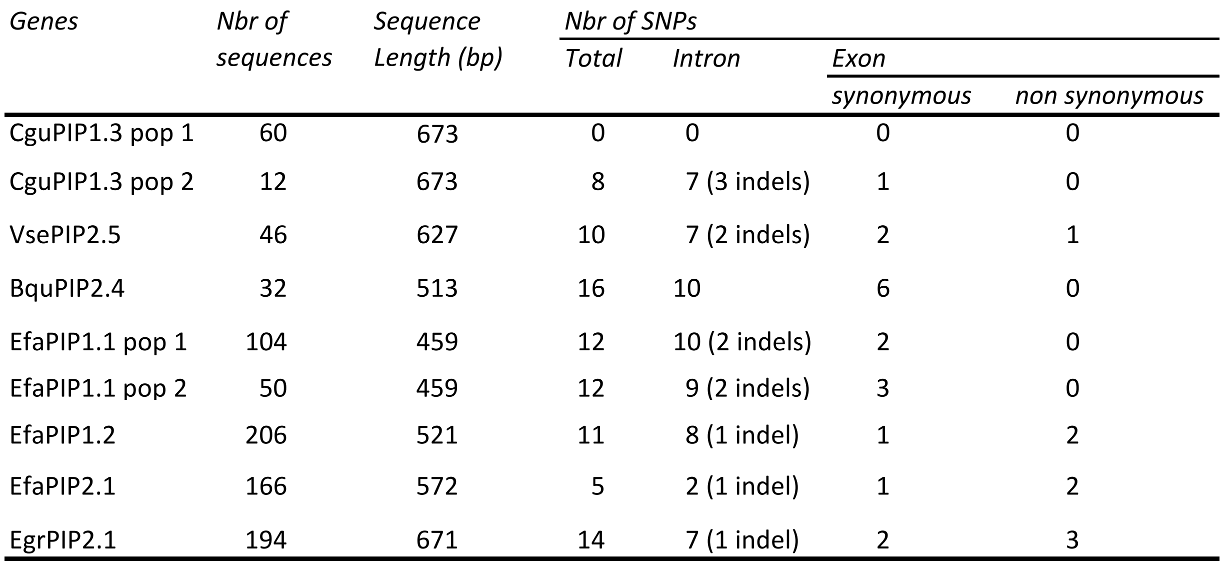
[2] For *E.grandiflora* , the same primers and conditions were used for PCR amplifications.

Species	***Name***	***Product Size (bp)***	***Primer length (bp)***		***Primer sequence 5' -> 3'***	***tm***	***Amplification (anneling at 59°C)***	***Comments***
Bombacopsis	*BquPIP2.1+0017*	*657*	*20*	*F*	*GCCGGTATCTCTGGTGAGTG*	*60.68*	*yes*	

	BquPIP2.1-0672		*20*	*R*	*CCACGCCTTCTCTTTGTTGT*	*60.29*	🟩
Bombacopsis	*BquPIP2.2+0029*	*601*	*20*	*F*	*GGTTGTTTTTGGCCCGTAAG*	*61.59*	🟥 *no*
	BquPIP2.2-0629		*19*	*R*	*CGACCCAGAAGACCCAGTG*	*61.70*	

Table 2. Polymorphism description of each gene region. Gene name correspond to genus and specie first letters followed by the name of BLAST results. The hyphen means that there is no polymorphism.

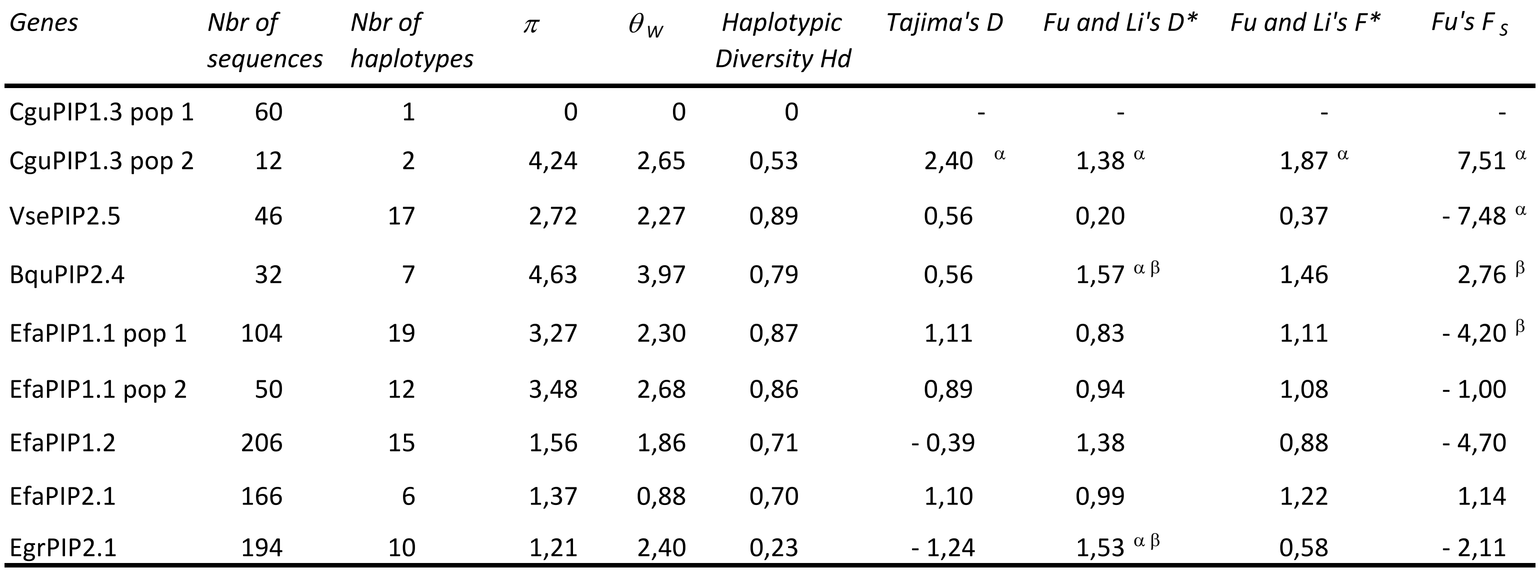
Commentaire [IS32]: **DELPHINE:** This table appears here as a figure, and needs to be back-converted into a table.



Genes	*Nbr of sequences*	*Sequence Length (bp)*	*Nbr of SNPs*			
			Total	*Intron*	*Exon*	
					synonymous	*non synonymous*
CguPIP1.3 pop 1	60	673	0	0	0	0
CguPIP1.3 pop 2	12	673	8	7 (3 indels)	1	0
VsePIP2.5	46	627	10	7 (2 indels)	2	1
BquPIP2.4	32	513	16	10	6	0
EfaPIP1.1 pop 1	104	459	12	10 (2 indels)	2	0
EfaPIP1.1 pop 2	50	459	12	9 (2 indels)	3	0
EfaPIP1.2	206	521	11	8 (1 indel)	1	2
EfaPIP2.1	166	572	5	2 (1 indel)	1	2
EgrPIP2.1	194	671	14	7 (1 indel)	2	3

Table 3. Genetic diversity and results of neutrality tests for each gene. Analyses were performed with and without taking recombination into account. Test significance is indicated by $^{\alpha}$ for P-values < 5% without recombination and by $^{\beta}$ for P-values < 5% with recombination.

Commentaire [IS33]: **DELPHINE:** This table appears here as a figure, and needs to be back-converted into a table.
Moreover: only values with recombination should be displayed, as it is anyway the most precise way to analyse data.



Genes	*Nbr of sequences*	*Nbr of haplotypes*	π	θ_W	*Haplotypic Diversity Hd*	*Tajima's D*	*Fu and Li's D**	*Fu and Li's F**	*Fu's F_S*
CguPIP1.3 pop 1	60	1	0	0	0	-	-	-	-
CguPIP1.3 pop 2	12	2	4,24	2,65	0,53	2,40 $^{\alpha}$	1,38 $^{\alpha}$	1,87 $^{\alpha}$	7,51 $^{\alpha}$
VsePIP2.5	46	17	2,72	2,27	0,89	0,56	0,20	0,37	- 7,48 $^{\alpha}$
BquPIP2.4	32	7	4,63	3,97	0,79	0,56	1,57 $^{\alpha\,\beta}$	1,46	2,76 $^{\beta}$
EfaPIP1.1 pop 1	104	19	3,27	2,30	0,87	1,11	0,83	1,11	- 4,20 $^{\beta}$
EfaPIP1.1 pop 2	50	12	3,48	2,68	0,86	0,89	0,94	1,08	- 1,00
EfaPIP1.2	206	15	1,56	1,86	0,71	- 0,39	1,38	0,88	- 4,70
EfaPIP2.1	166	6	1,37	0,88	0,70	1,10	0,99	1,22	1,14
EgrPIP2.1	194	10	1,21	2,40	0,23	- 1,24	1,53 $^{\alpha\,\beta}$	0,58	- 2,11

Table 4. Results of tests for demographic changes based on chloroplast SSR loci. Ccmp1 through ccmp10: names of chloroplast SSR loci. The value of DH/sd, standardised departure from heterozygosity expected under mutation/drift equilibrium, is reported. Positive values indicate population contraction, negative values indicate population expansion. Signficance levels: *, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001. N = number of informative loci; e = number of loci indicating population expansion; c = number of loci indicating population contraction; m = monomorphic locus; na = locus which did not produce any PCR product.

Population	N (e/c)	ccmp1	ccmp2	ccmp3	ccmp4	ccmp5	ccmp6	ccmp7	ccmp10
C. guianensis Pop 1	7 (2/0)	na	-4.187**	-2.840*	0.886	-1.062	-1.388	-1.049	-0.908
C. guianensis Pop 2	5 (0/0)	na	0.010	m	-1.274	m	0.424	0.378	0.353
V. sebifera	1 (0/0)	m	m	-0.517	m	m	m	m	m
B. quinata	1 (0/0)	m	m	m	m	na	0.464	na	m
E. falcata Pop 1	6 (3/0)	m	-0.506	-3.109**	-1.340	na	-2.159*	-2.084*	-1.291
E. falcata Pop 2	3 (1/0)	m	-0.243	m	0.136	na	m	-2.969*	m
E. grandiflora	5 (3/0)	-2.925**	na	-1.288	0.941	na	-3.054**	-4.212**	na

Captions to figures

Figure 1.The two PIP1 – PIP2 conserved regions chosen based on ClustalW alignments (only a subsample of the aligned sequences is shown). Blue rectangles represent transmembrane helices two and six, in which the universal primers were chosen (amino-acid sequences in red).

Figure 2. Representation of the different fragments localization compared to *Arabidopsis thaliana* PIP structure. Blue rectangles, located on the diagram of the Arabidopsis gene structure, represent the six transmembrane helices.

Commentaire [IS34]:
DELPHINE: is this a part of a published figure?

Figure 3. Sliding window of Fu and Li's *F*- and *D*-tests and Tajima's *D*, along EfaPIP1.1. For 3 sites (286, 287 and 288) all statistics are negative and significantly different from zero.

Commentaire [IS35]:
DELPHINE: this result is very interesting (see what I have added to the results section) – have you tested sliding windows for the other genes?
The start and end of exon and intron (as deduced from alignments with cDNAs) should be indicated in the figure (for EfPIP1.1, the exon fragment ends at base 180)

Figure 1.

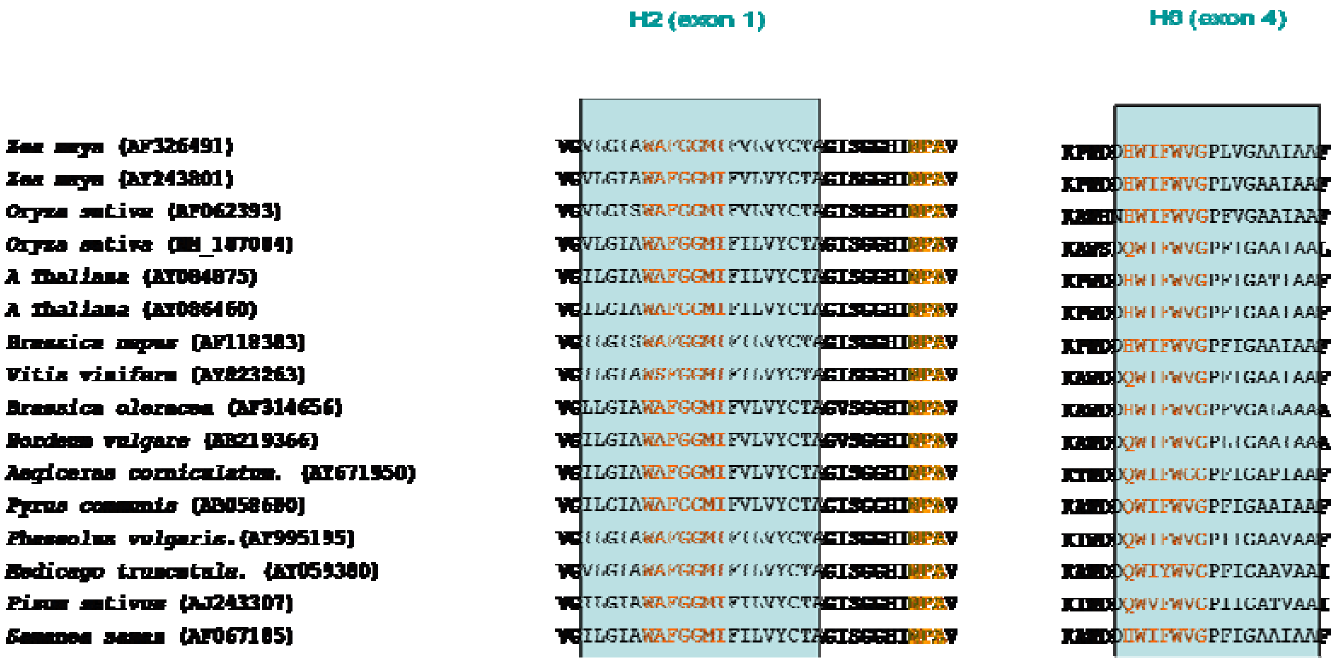

H2 (exon 1)
H8 (exon 4)
Zea mays (AF326491) VGVLGIAWAFGGMIFVLVYCTAGISGGHINPAV KPWDDHWIFWVGPLVGAAIAAF
Zea mays (AY243801) VGVLGIAWAFGGMIFVLVYCTAGISGGHINPAV KPWDDHWIFWVGPLVGAAIAAF
Oryza sativa (AF062393) VGVLGISWAFGGMIFVLVYCTAGISGGHINPAV KAWHNHWIFWVGPFVGAAIAAF
Oryza sativa (NM_187084) VGVLGIAWAFGGMIFILVYCTAGISGGHINPAV KAWSDQWIFWVGPFIGAAIAAL
A Thaliana (AY084875) VGILGIAWAFGGMIFILVYCTAGISGGHINPAV KPWDDHWIFWVGPFIGATIAAF
A Thaliana (AY086460) VGILGIAWAFGGMIFILVYCTAGISGGHINPAV KPWDDHWIFWVGPFIGAAIAAF
Brassica napus (AF118383) VGILGISWAFGGMIFILVYCTAGISGGHINPAV KPWDDHWIFWVGPFIGAAIAAF
Vitis vinifera (AY823263) VGILGIAWSFGGMIFILVYCTAGISGGHINPAV KAWDDQWIFWVGPFIGAAIAAF
Brassica oleracea (AF314656) VGLLGIAWAFGGMIFVLVYCTAGVSGGHINPAV KAWDDHWIFWVGPFVGALAAAA
Hordeum vulgare (AB219366) VGILGIAWAFGGMIFVLVYCTAGVSGGHINPAV KAWDDQWIFWVGPLIGAAIAAA
Aegiceras corniculatum. (AY671950) VGILGIAWAFGGMIFVLVYCTAGISGGHINPAV KTWDDQWIFWCGPFIGAPIAAF
Pyrus communis (AB058680) VGILGIAWAFGGMIFVLVYCTAGISGGHINPAV KAWDDQWIFWVGPFIGAAIAAF
Phaseolus vulgaris.(AY995195) VGILGIAWAFGGMIFILVYCTAGISGGHINPAV KTWDDQWIFWVGPIIGAAVAAF
Medicago truncatula. (AY059380) VGVLGIAWAFGGMIFILVYCTAGISGGHINPAV KAWDDQWIYWVGPFIGAAVAAI
Pisum sativum (AJ243307) VGILGIAWAFGGMIFILVYCTAGISGGHINPAV KTWDDQWVFWVGPIIGATVAAI
Samanea saman (AF067185) VGILGIAWAFGGMIFILVYCTAGISGGHINPAV KAWDDHWIFWVGPFIGAAIAAF

Figure 2

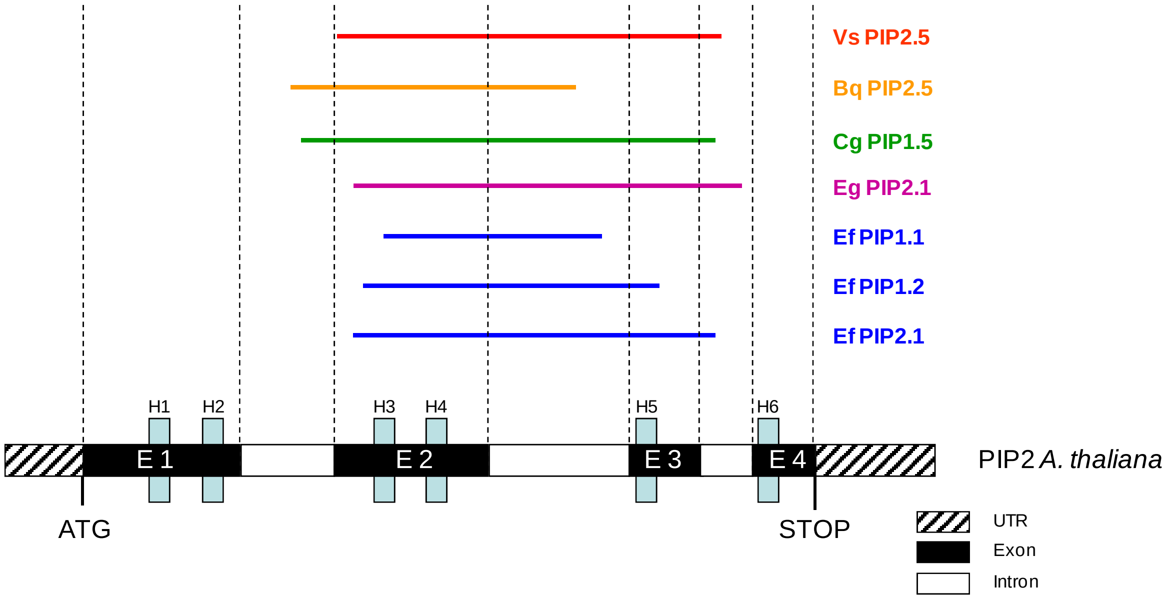
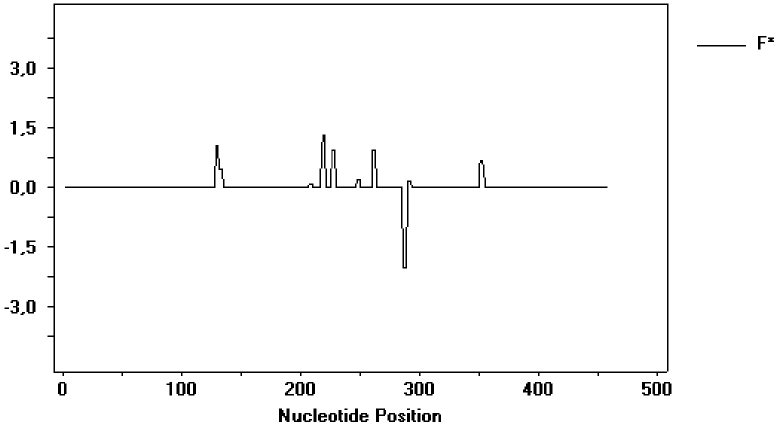
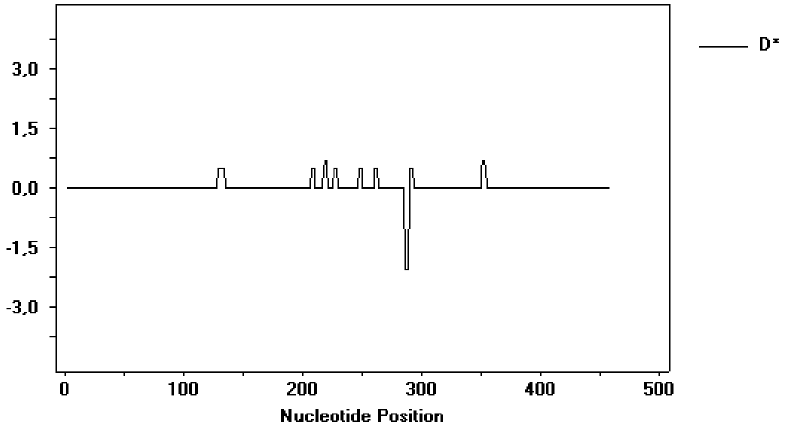
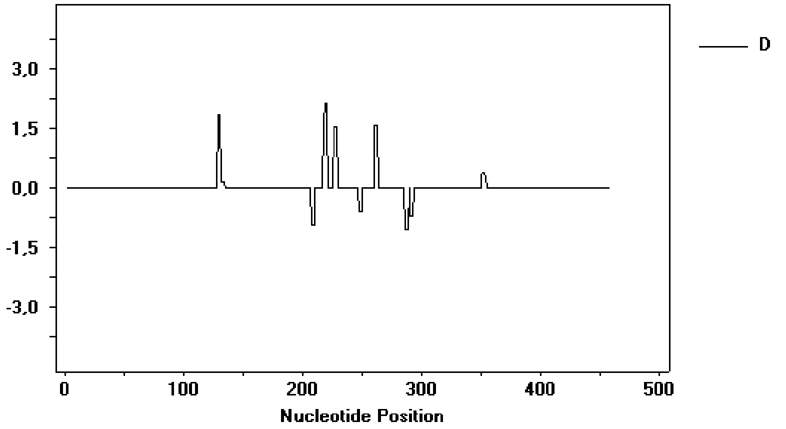

Vs PIP2.5
Bq PIP2.5
Cg PIP1.5
Eg PIP2.1
Ef PIP1.1
Ef PIP1.2
Ef PIP2.1
H1
H2
H3
H4
H5
H6
E 1
E 2
E 3
E 4
PIP2 *A. thaliana*
ATG
STOP
UTR
Exon
Intron

Figure 3.




D*
3,0
1,5
0,0
-1,5
-3,0
0
100
200
300
400
500
Nucleotide Position


D
3,0
1,5
0,0
-1,5
-3,0
0
100
200
300
400
500
Nucleotide Position