



Interaction between gene flow and selection in hybridizing oak species

Garnier-Géré Pauline

Abadie Pierre, Navascues Miguel, Lang Tiange, Brendel Oliver, Scotti-Saintagne Caroline, Gerber Sophie, Bodénès Catherine, Petit Rémy, Kremer Antoine



UMR 7625,
Ecology &
Evolution Lab



UPMC
PARIS UNIVERSITAS

Cestas

B i o G e C o



Biodiversité, gènes & communautés

Champenoux

UMR EEF INRA-UHP

1137, Forestry Ecology
& Ecophysiology

Contributions

***White oak species
mapping, population
genetics, and evolution***

Sophie Gerber

Catherine Bodénès

Caroline Scotti-Saintagne

Remy Petit

Antoine Kremer

***Water Use efficiency in oaks &
ERECTA gene homolog***

Oliver Brendel

***Coalescent method for testing
introgression versus isolation***

Miguel de Navascues

***Interspecific cross-compatibility –
Molecular evolution Genes in RI***

Pierre Abadie, PhD

POSTER

***Oak allelic resequencing project –
bioinformatic tools development for
automatic analyses***

Tiange Lang, Post-doc

POSTER

Speciation genetics: evolving approaches

Shift from **whole-genome view of speciation** (BSC and primacy of geographical RI) towards an alternative **genic view of speciation** (*Wu & Ting 2004, Noor & Feder 2006*)

→ In plants and animals, **adaptive divergence and speciation can occur in the presence of gene flow**

→ Increasing examples showing **genomic islands** of differentiation & speciation (*Turner et al 2005, Harr 2006, Savolainen et al 2009*)

Advances in genomic techniques → understanding of processes, based on molecular evolution inferences of potentially important genes

→ « Speciation » genes? Very different functions, not only in reproductive isolation

Adaptation to soil pH in exotic palms

Host-pathogens interactions

Unknown functions

Genes causing hybrid inviability or sterility

Introduction

The Wu's model of species differentiation (2001,2004)

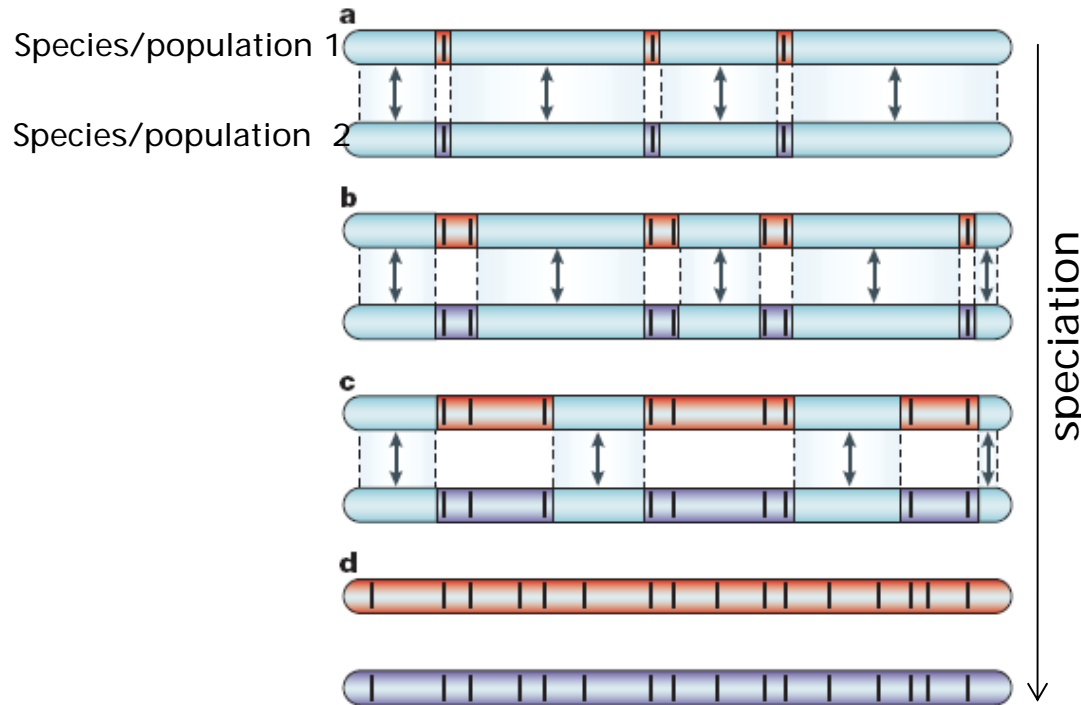


Figure 1 | The genic view of species differentiation.

→ Mosaic of regions

With low differentiation,
« permeable » to **gene flow**



With higher divergence → genes
submitted to **selection** (RI or
ecological differentiation)

→ When & if the differentiated
regions expand, speciation process
lead to a complete isolation of
genomes

which mechanisms favor this process?

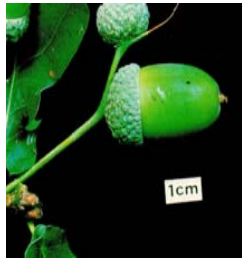
Which genes, which biological functions?

Equilibrium before complete isolation?

The complex of white oak species *Quercus robur* & *Q. petraea*

Sessile & Pedunculate white oaks

Despite acknowledged (if not large) morphological and ecological differentiations, very low molecular differentiation (Nuclear Markers & CpDNA) in « most » parts of the genome BUT several « clusters » of higher F_{st} values



Pedunculate oak



Sessile oak



Working Hypothesis: strong **diversifying selection** despite **high gene flow** producing such a mosaic of genomic regions more or less « permeable » to gene flow, with « hot-spots » of differentiation between species.

The complex of white oak species *Quercus robur* & *Q. petraea*

Summary of 2 decades of discoveries (> 30 scientific papers)

Botanists & ecologists can distinguish the species → ecological species concept

The European Jay too (shape and position of acorns)!

Natural Selection inferred

However, very low molecular differentiation at nuclear markers & CpDNA sharing

→ Evidence of **recurrent hybridization** (*Lepais et al. 2009*, up to 30% of genotypes with hybrid or backcross status using SSRs & genetic assignments methods) & demographic component of hybridization and introgression direction

Gene flow demonstrated

Genome scan of differentiation showing non-random clusters of high F_{st} values on 289 markers (*Scotti-Saintagne et al. 2004*)

→ Selection inferred

The complex of white oak species *Quercus robur* & *Q. petraea*

Hybridization as a mechanism of invasion in oaks

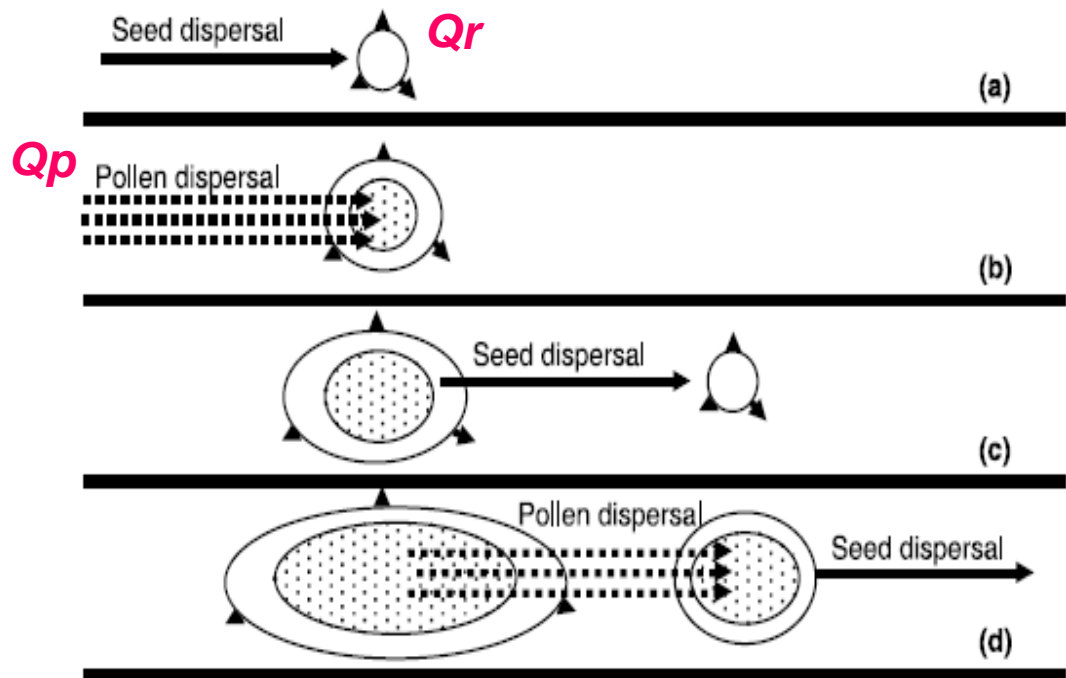
Colonization dynamics

Rémy J. Petit, Catherine Bodénès, Alexis Ducouso, Guy Roussel and Antoine Kremer

Better dispersal abilities of *Q. robur* for both seed and pollen (but pollen flow more efficient in both species) → *Q. robur* pioneer species in colonization dynamics of new sites

Crossings easier from male *petraea* to female *robur* → **asymmetric hybridization**

Model of invasion by *Q. petraea* through hybridization, of the range occupied by the pioneer *Q. robur* species,



The complex of white oak species *Quercus robur* & *Q. petraea*

Is it that simple? --> For the system to work, *Q. petraea* when it first hybridizes, need to hybridize again, fairly efficiently with pollen from outsider same species (long distance pollen flow)

→ Hypothesis of an asymmetric barrier (Petit et al. 2003)

Where, in the process of invasion by *Q. petraea*, there would be a **reinforcement** of reproductive barriers from *Q. robur* to *Q. petraea* to allow *Q. petraea* to produce backcross genotypes and regenerate

→ so even though hybrids should be less fit, hybrid fitness needs to be high enough so that they can survive and reproduce

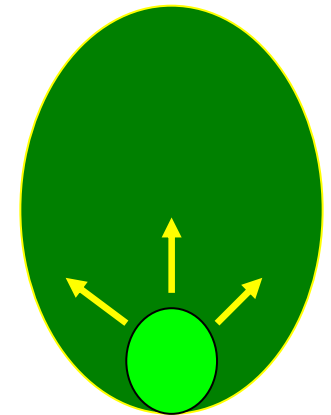
→ Selective pressures to evolve an incompatibility barrier will be especially strong in the species which is at a numerical disadvantage early in succession, *Q. petraea*.

→ Many hypotheses to be confirmed and unanswered questions:

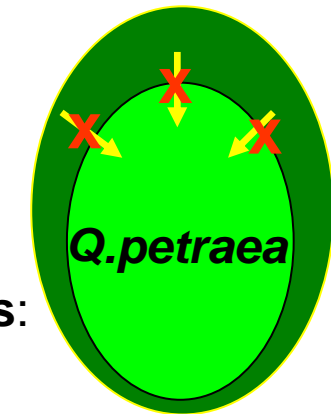
Genes involved in reproductive isolation in both species?

Intensity of selection at those genes.

→ Contribution of sequence data from the oak resequencing project?



Pollen Flow



Pollen Flow



The Oak allelic resequencing project



Funding: EVOLTREE, ANR Transbiodiv, ANR Realtime, Linktree

Objectives

Better characterize *Q. petraea* and *Q. robur* hybridizing species genome nucleotide diversity

- *Focus on differentiated regions among species,*
- *compare genes from # biological functions potentially involved in adaptive divergence,*
- *but also assess shared diversity between species,*
- *estimate distribution for different summary statistics, **population genomics methods...***

Approach

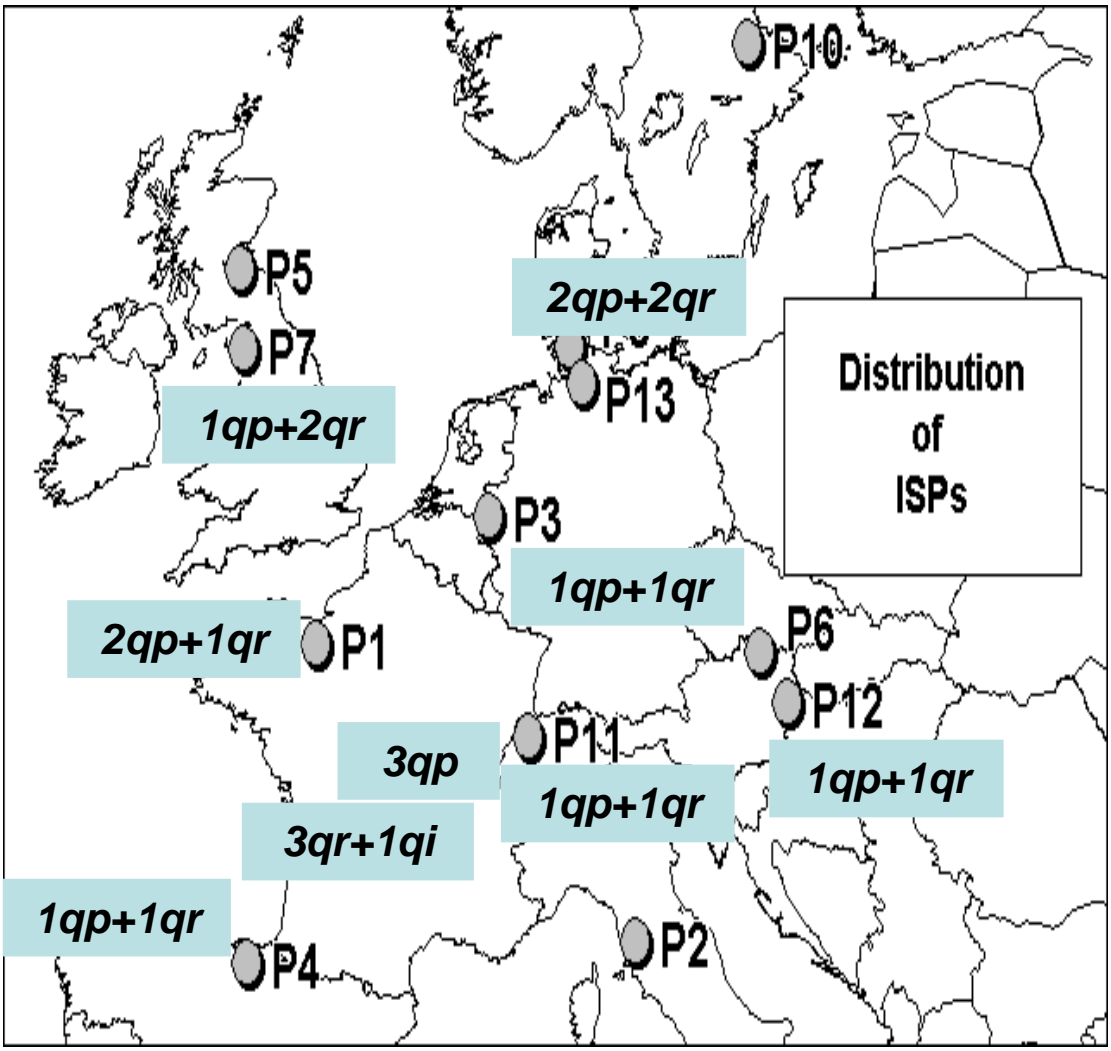
- Amplification & resequencing of 1000 amplicons
- Bioinformatics & sequence data production strategy
- Development of automatic pipeline of analyses

POSTER
Lang et al.

Oak allelic resequencing project:

Genotypes

{ 12 *Quercus robur*
 12 *Q. Petraea*
 1 *Q. Ilex*
 sampled across the
 natural distribution
 (OakFlow project ISPs)



Choice of gene fragments → different lists / different objectives

Functional & expressional Candidate genes

→ various functional traits important for **ecological differentiation** and/or **reproductive isolation** (drought stress tolerance, Anoxia, pollen/pistil interaction, crossing compatibility, bud phenology, host-pathogens interactions)

Reference (random) set of amplicons

→ get « background /experimental distribution » for various statistics to compare to particular/targeted genes

→ compare patterns in both species for a particular set of genes to this reference

Oak allelic resequencing project: Sequence data summary (1)

Large delays → Last 30% of data on 29/09/09!

982 amplicons were sequenced for **786 genes**

a **good** sequence = one with 50% good quality nucleotides (phd score 30)

➤ If successful amplicon = one with at least 24 **good** sequences (max=48)

➤ If successful amplicon = one with at least 12 **good** sequences (max=48)

➤ *Q. ilex* reference species was amplified for ~55 % of amplicons

→ **Importance of control of data quality at every step of sequence production from primer design → last step of resequencing on more individuals** (~50% global success rate on 2000 amplicons (1968 effective) – 2 individuals – phd score 40)

Success rates

Amplicon	Gene
64%	67%
85%	86%

POSTER
Lang et al.

Oak allelic resequencing project: Sequence data summary (2)

- ~50% of amplicons have introns of size > 100 bp
- Average length for 837 successful fragments: **563 bp** (min: 176 max:1390)
- More detailed analyses **on 190 genes** → ~1500 SNPs in both species (with 4% monomorphic genes in both species)
- → Potential of **6000 SNPs** on 786 genes → larger-scale genotyping

Overall, NO allelic substitution at SNPs were observed among species, majority of shared polymorphisms and less often exclusive sites → status of hybridizing species

Oak allelic resequencing project → Preliminary results on nucleotide diversity and differentiation (190 genes) (1)

Expressionnal & functional candidates from literature

Functional candidates from different gene Databases in model & related species, based on keywords from metabolic pathways

Of resistance mechanisms (general & more specific)

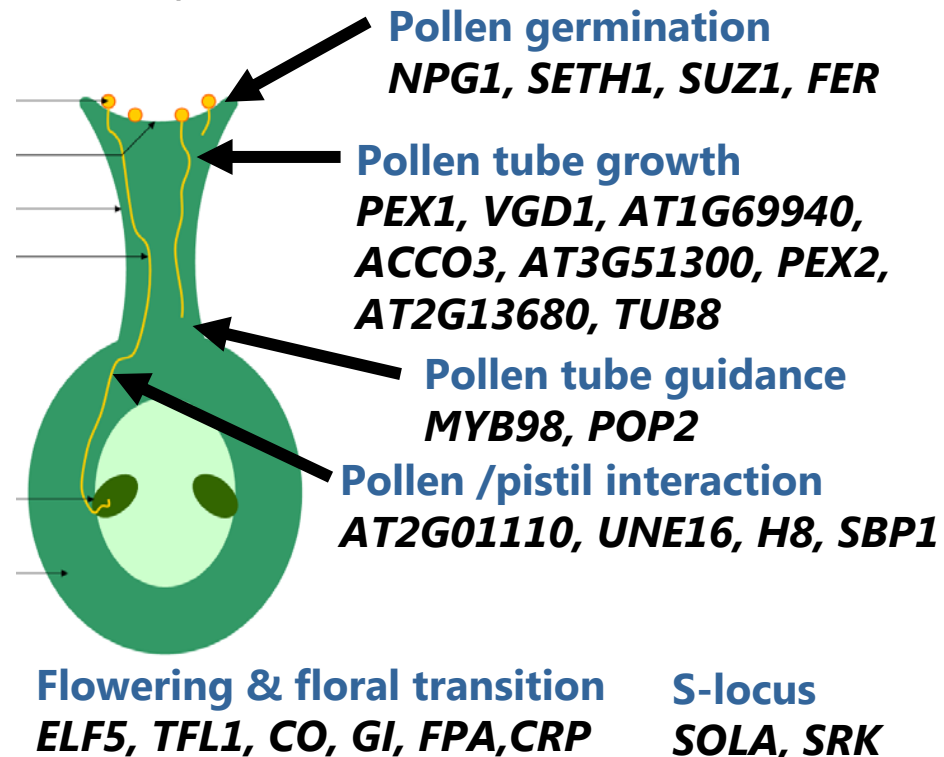
➤ **Insect-plants interactions/ herbivory genes (38)**

➤ **Host-Fungal pathogens interactions genes (42)**

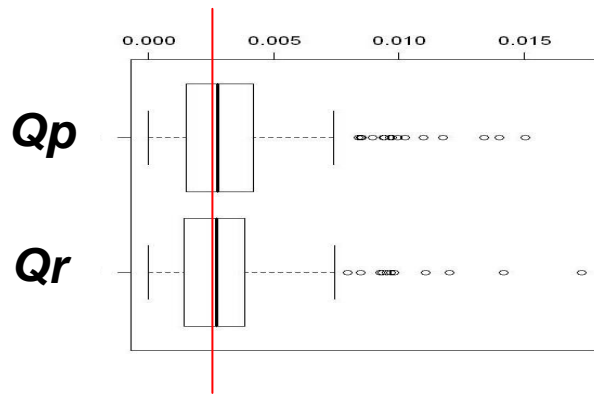
+ **Reference genes (82)** < random draw of ESTs in oak assembly

➤ **Reproductive barriers genes (28)**

Oak female flower during fecundation



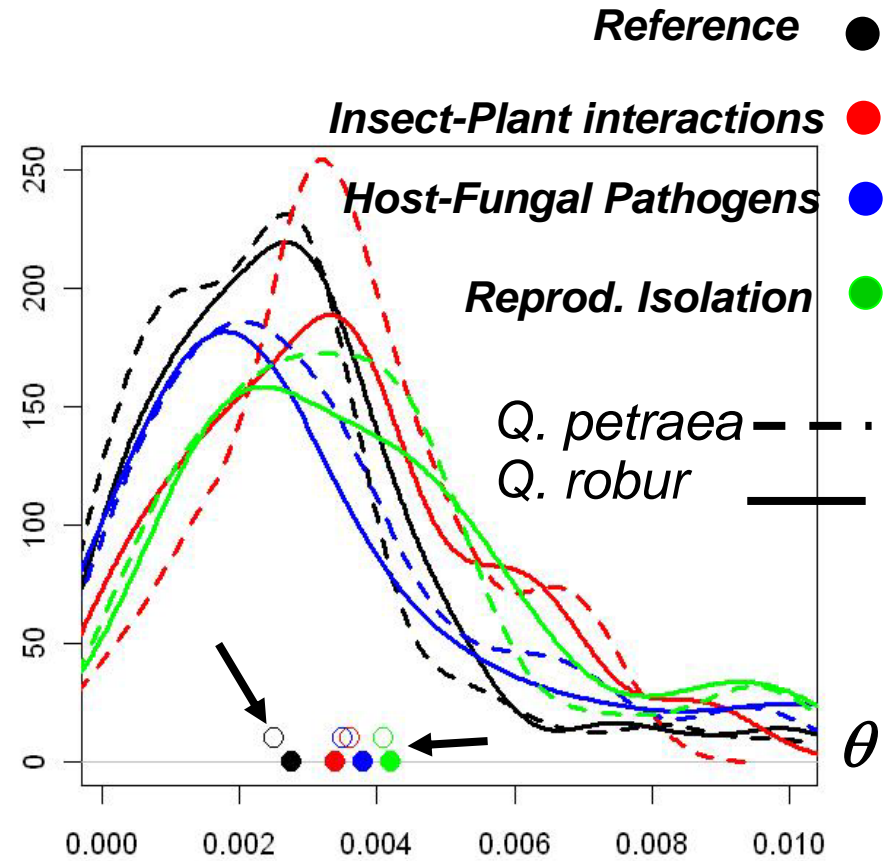
Oak allelic resequencing project → Preliminary results on nucleotide diversity (190 genes) (2)



Mean π_i of Qp vs Qr → ~0.003 for both species on 162 fragments

Same for all the lists but depends on genes

→ Slightly greater diversity in *Q. petraea* less obvious than on neutral markers

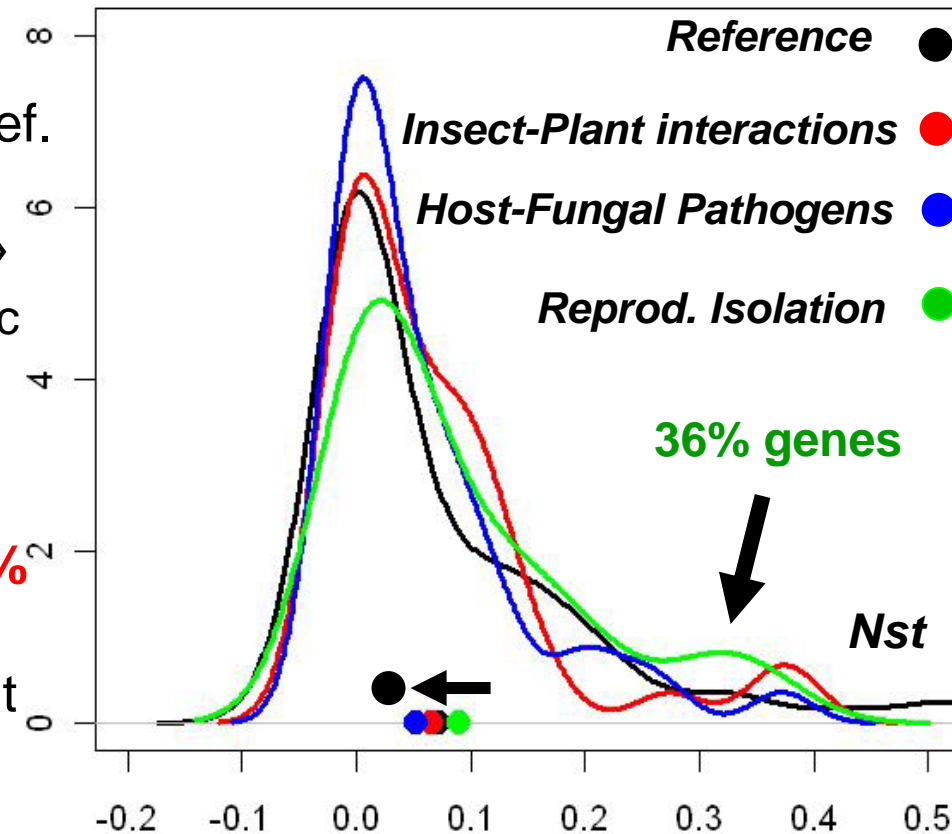


Large range of diversity - 10 fold variation among genes in the different gene lists

Oak allelic resequencing project → Preliminary results on nucleotide differentiation *Nst* (190 genes) (3)

- Skewed distribution of *Nst* values
- Average nucleotide differentiation (*Nst*) across genes: **7.0%** (7.3 % for ref. list) but **median=3%** (2.4%)
- Consistent with previous « neutral » estimates of ~3% < EST-based genomic sequences
- Range: from **0 to 53%**, with **26% of genes with *Nst* values > 10%**
- Differentiated regions more frequent than previously thought...those including genes

Comparison among genes lists



Future objective: → model-based genome scan of differentiation to identify outliers for positive (diversifying) selection

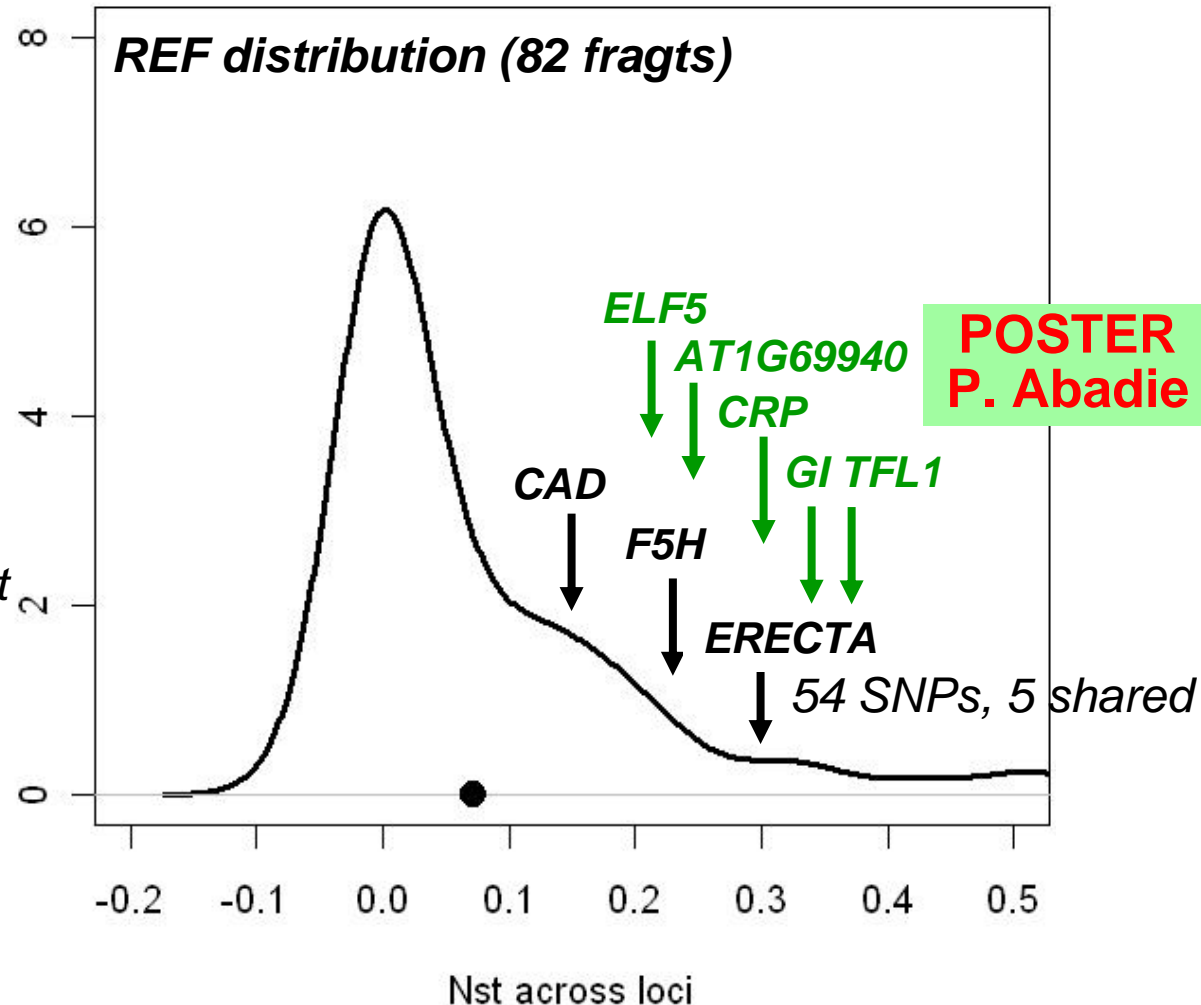
A few interesting candidates for species divergence? – insights into biological functions (*genetic basis of adaptive divergence*)

Funding: French Bureau of Genetic Resources, Evoltree, Transbiodiv

ERECTA: linked to water use efficiency (stomatal density) with **CAD**, **F5H** (drought tol.)

RI genes: involved in pollen-pistil interaction, Flowering pathway (4)

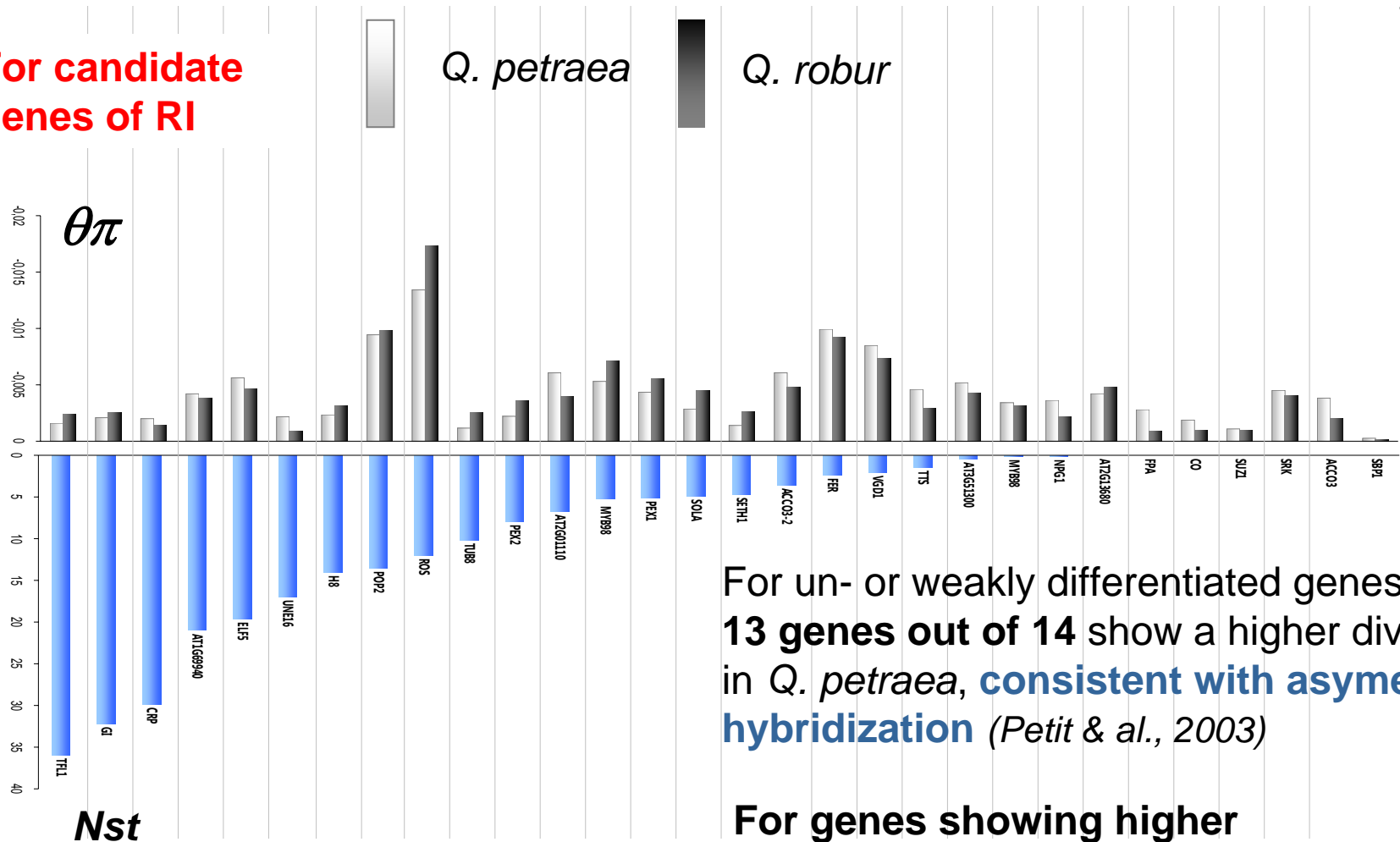
REF genes: 4 genes with *Nst* values from 30% to 53%
→ **UNKNOWN functions!!**



Oaks nucleotide diversity & diff study → insights into processes (1)

→ Distinguish between genes showing **low** differentiation versus **higher** differentiation which could be the consequence of selection

For candidate genes of RI



For un- or weakly differentiated genes, **13 genes out of 14** show a higher diversity in *Q. petraea*, **consistent with asymmetric hybridization** (Petit & al., 2003)

For genes showing higher differentiation, **11/16** show the contrary

Oaks nucleotide diversity & diff study → insights into processes (2)

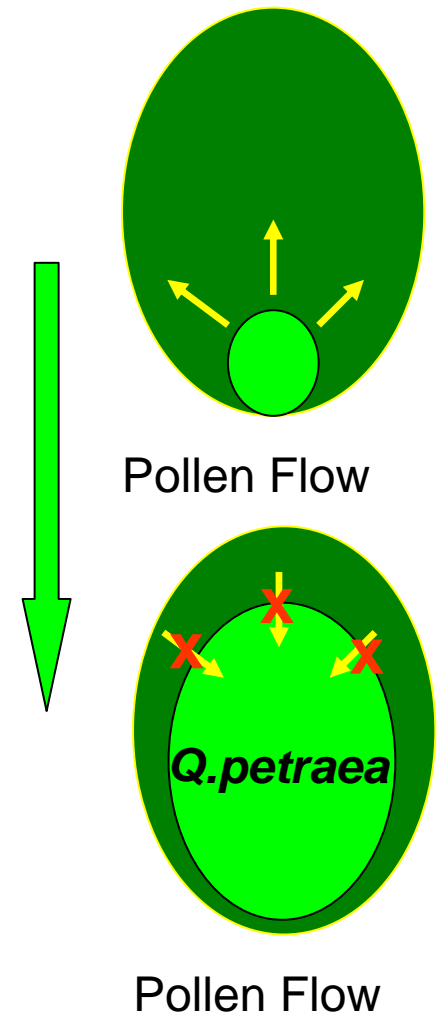
What if for genes where there would be stronger selection due to divergent selection (→ more differentiated among species)...

So what if for those genes specifically involved in RI, the selection was stronger in *Q. petraea*? < trend towards lower diversity at those genes

→ This would fit the model of invasion of *Q. robur* by *Q. petraea* by asymmetric hybridization

Those stronger barriers are necessary for *Q. petraea* hybrids to avoid being invaded back by *Q. robur* local pollen.

→ So for genes potentially involved in RI, we may have some preliminary results illustrating the process of selection on reproductive barriers which is necessary for *Q. petraea* to be colonizing new areas through hybridization with *Q. robur*.



650 genes on up to 25 individuals

→ we have some population genomics data!!

- *Estimate F_{st}/N_{st} at both SNP & haplotype level (< haplotype reconstruction), Compare to EST Dbase screening*
 - *methodological issue on which differentiation statistics is best*
- *Distribution of nucleotide diversity and other statistics that would allow to detect selection signatures, compare among different types of genes and the fragments used as random reference → Tests integrating demography*
- *Test of outlier genes for different summary statistics linked to differentiation*
- *Genetic basis of speciation → relative roles of flowering pathway genes and other genes not involved directly in Reproductive Isolation*
- *Variability in K_s ratios → inferences for timing since divergence / introgression & extent/recurrence of gene flow*

Acknowledgments

ESTs & genomic resources

C. Bodénès, F. Salin
V. Léger, G. Le Provost, C. Sulmon
F. Hubert

Oak allelic resequencing project contributors

J-M Frigerio
M-L Desprez-Loustau
T. Decourcelle
C. Robin, C. Burban
G. Le Provost, E. Guichoux
M-P Reviron, C. Plomion

Oak Material management and crossing/Field experiments

G. Roussel
E. Dreyer
J-M Louvet
Alexis Ducousso

Funding



Contract number
016322



ANR 06-BDIV-003-04
TRANSBIODIV
ANR REALTIME
N°59000256

French Ministry of education and
research



French Bureau of
Genetic Resources