



HAL
open science

Intérêts d’approches génomiques complémentaires pour l’analyse de la diversité génétique fonctionnelle de populations naturelles de Peuplier Noir

Catherine Bastien, Véronique Jorge, Vincent Segura, Christelle Aluome, Vanina Guérin, Redouane El Malki, Pauline Sandra Paulstephenraj, Marie-Christine Le Paslier, Aurélie A. Berard, Aurelie A. Chauveau, et al.

► To cite this version:

Catherine Bastien, Véronique Jorge, Vincent Segura, Christelle Aluome, Vanina Guérin, et al.. Intérêts d’approches génomiques complémentaires pour l’analyse de la diversité génétique fonctionnelle de populations naturelles de Peuplier Noir. Colloque EPGV 2014 "Détection, Gestion et Analyse du Polymorphisme des Génomes Végétaux", Jun 2014, Evry, France. 17 p. <hal-02801095>

HAL Id: hal-02801095

<https://hal.inrae.fr/hal-02801095v1>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



Intérêts d'approches génomiques complémentaires pour l'analyse de la diversité génétique fonctionnelle de populations naturelles de Peuplier Noir

C. Bastien¹, V. Jorge¹, V. Segura¹, C. Aluome³, V. Guérin¹, R. El Malki¹, P. PaulStephenRaj², MC. Le Paslier³, A. Berard³, A. Chauveau³, R. Bounon³, D. Brunel³, M. Villar¹, P. Faivre-Rampant²



IUFRO June 2014

¹INRA UR 0588 AGPF, Centre INRA Val de Loire, 2163 avenue de la Pomme de Pin, CS 40001 – Ardon 45075 Orléans, France

²INRA UMR 1165 URGV, 2 rue Gaston Crémieux, 91057 Evry-France

³INRA US 1279 EPGV/CEA/CNG, 2 rue Gaston Crémieux, 91057 Evry, France

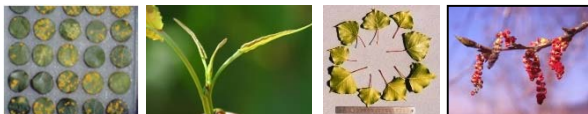
Des études génétiques centrées sur *Populus nigra* L.

Diversité en populations naturelles

- Espèce majeure des ripisylves
- Une aire de distribution vaste mais des menaces localement



- Une variabilité importante pour différents traits adaptatifs



Diversité pour les besoins en sélection

P. deltoides X *P. nigra*



- Hétérosis en croisement interspécifique
- Variétés clonales



Flux de gènes et introgression

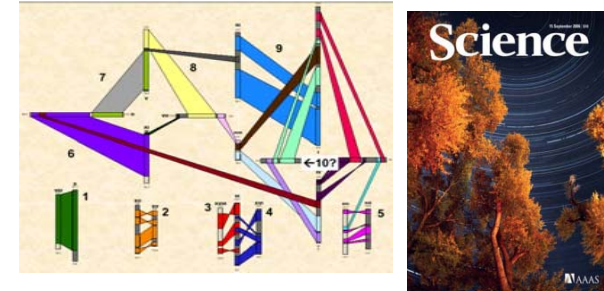
Les ressources génomiques disponibles

- Un génome de référence annoté disponible depuis 2006 pour une espèce proche *P. trichocarpa*

n= 19, taille= 480 Mb, 200kb/cM

Des duplications récentes

Un DL qui décroît très rapidement



- Reséquençages d'EST sur cultivars hybrides
- Banque BAC sur *P. trichocarpa* '101-74'

➔ Des développements de SNP à centrer sur l'espèce *P. nigra* et des génotypages pour différents objectifs

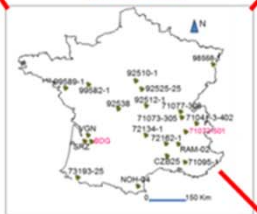
La stratégie choisie



Population d'association
1100 individus



Panel de découverte de SNP
21 individus



Plan factoriel de croisement 5x4 *P. nigra*
937 individus

	SRZ	BDG	71077-308	92510-1
VGN	33	40	40	40
CBZ25	40	39	32	39
71041-3-402		38	12	40
71072-501	40	315	40	
SSC	39	39	33	38

Pédigrée de cartographie
329 individus F1

Bien choisir les panels de diversité

- représenter la diversité génétique géographique et intra-population

1100

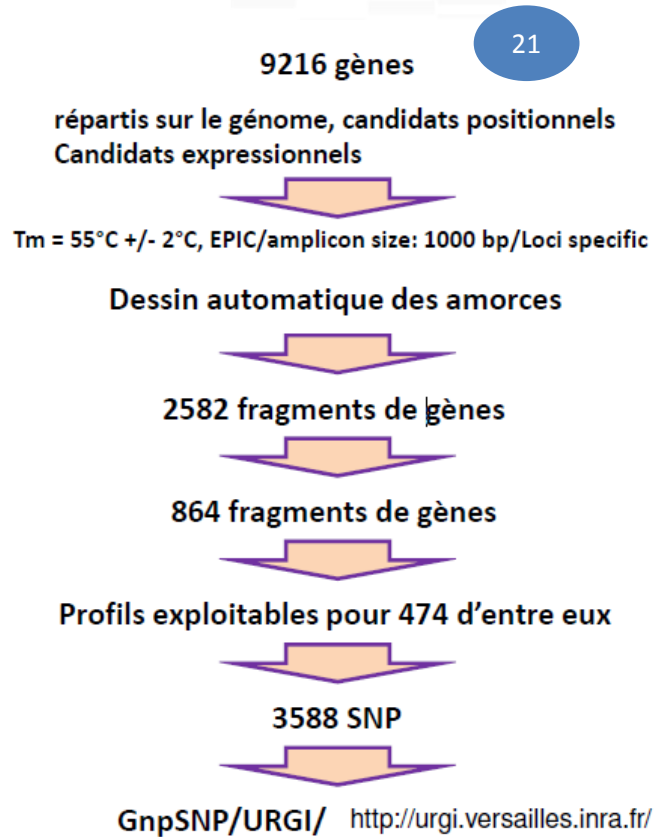
52

21

- Relier populations de cartographies génétiques ⁹, population d'association et population d'entraînement pour la sélection génomique ³⁰

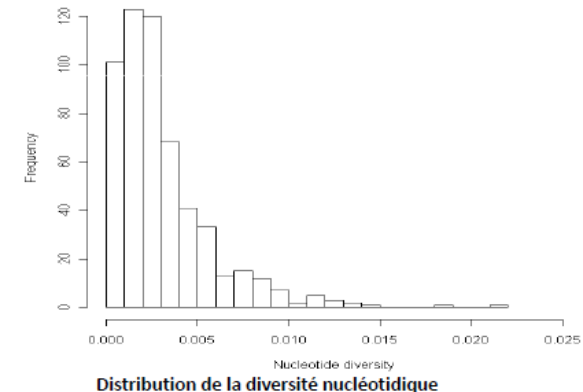
La stratégie choisie

(1) Leçons des re-séquençages SANGER



Alignement sur la séquence de référence
→ des SNP espèces spécifiques

Une diversité nucléotidique moyenne de **0,0037** comparable aux autres espèces du genre *Populus*



La stratégie choisie

(2) Du re-séquençage de génomes entiers

Un panel de 52 individus



GAll analyzer ou Hiseq 2000

Illumina Paired-End sample
protocol

Taille d'inserts : 300-600 bp

Runs pour 75, 100, 110 et 114 cycles

IGA

Udine, Italy

EPGV/IG-CNG

Evry, France

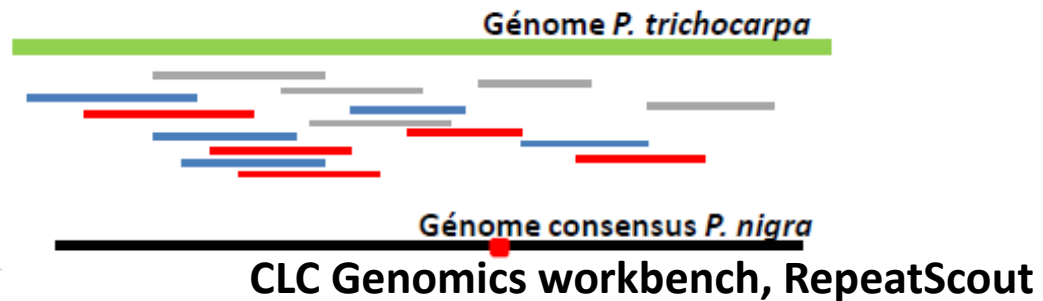
La stratégie choisie

(2) Du re-séquençage de génomes entiers (1)

4 individus *P. nigra*
Re-séquencés à couverture
élevée (>20X)



Obtention d'une
référence *P.nigra*



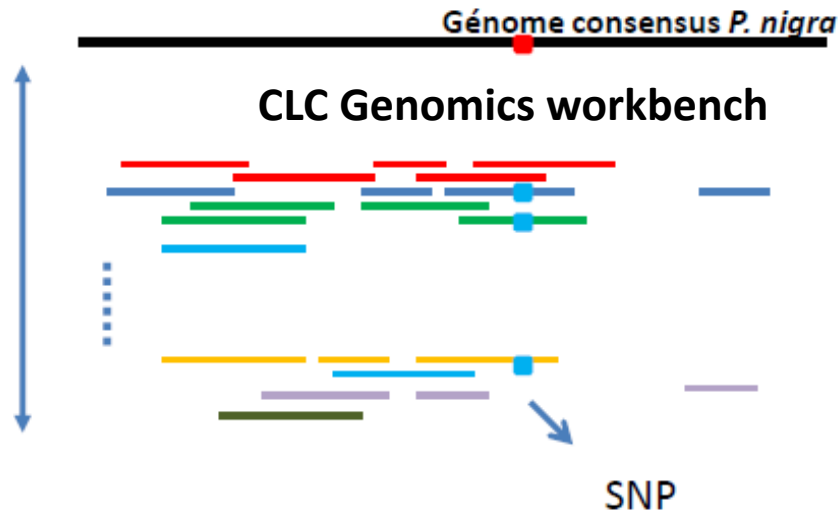
Clones	Gbp mapped	Reference covered (%)	Mean coverage (X)	Mean excluding zero coverage regions (X)
71077-308	~5.9	70	14.2	20.2
BDG	~8.4	77	19.9	25.8
BEN3	~7.1	79	16.9	21.5
POLI	~21.3	79	50.8	64.2

La stratégie choisie

(2) Du re-séquençage de génomes entiers (2)

Obtention d'une référence *P.nigra*

Alignement des séquences de 47 individus *P. nigra* re-séquencés à plus faible couverture ($2 < X < 10$)



- Maximum Coverage = 1.5 the average coverage
 - Minimum coverage = 0.1 to 0.5 the average coverage
 - MAF 35% for 20X, 15% for 2X
- (CLC Genomics workbench)



Genome Browser IGA



La stratégie choisie

(3) Des génotypages à la carte

- Détection d'introgression depuis *P. deltoides* dans les ressources naturelles *P. nigra* → 48-SNPlex Technology Applied Biosystems, EPGV-Evry, SNP Sanger, 0 SNP ds 60b flanquantes

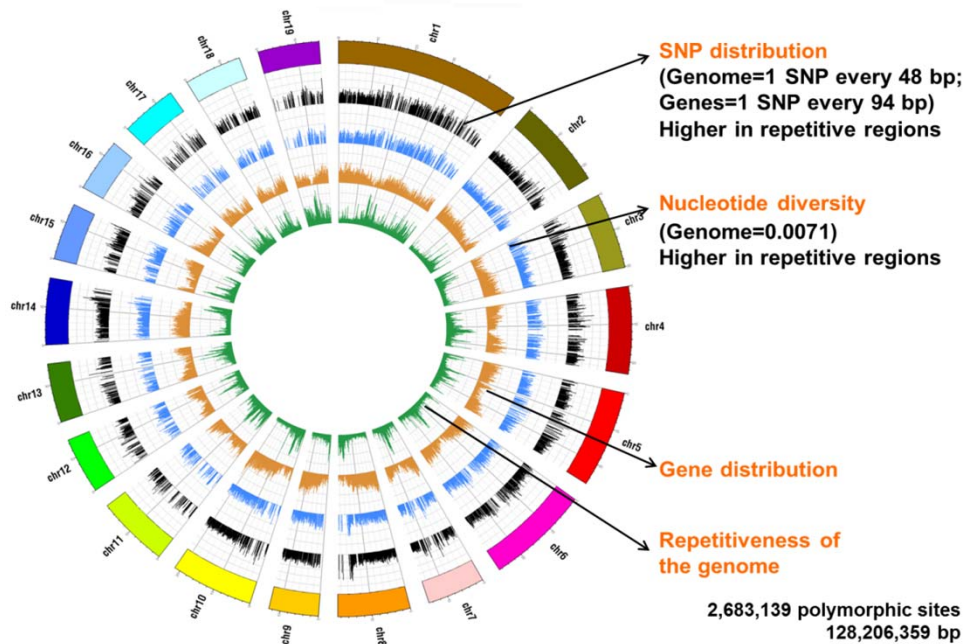
- Compléter un set de marqueurs microsatellites avec des SNP dans un pedigree de cartographie F1 → 96-SNPlex Technologie Applied Biosystems, EPGV-Evry, SNP Sanger, 0 SNP ds 60b flanquantes, unicité des loci

- Etudier la structuration génétique de la population d'association de 1100 individus → 1000 SNP
- Conduire de la génétique d'association dans des régions candidates
- Conduire des études d'association au sein d'un plan de croisement factoriel composé de 320 individus

Illumina Infinium
iSelect HD Custom
BeadChip ,
12000 beads
10331 SNP
EPGV-Evry

Résultats

Détection de SNP par re-séquençage de 52 génomes



SNP classification	# 1	#2	#3	3 clones
Total	~ 6.5 M	~ 6.5 M	~ 7.5 M	~ 7.3 M
<i>P.nigra-P.trichocarpa</i> SNPs	5,3 M	5,1 M	6,0 M	3,9 M
<i>P.nigra-P. nigra</i> SNPs	1,2 M	1,3 M	1,5 M	3,4 M
Genomic fraction covered	72%	75%	76%	80%

17 à 19% des SNP dans les gènes
81 à 83% en dehors des gènes

Validation de SNP NBS/Sanger: 96 164 sur 4 individus

Accuracy = ~100%

Specificity= 100%

False Discovery Rate = 7.6%

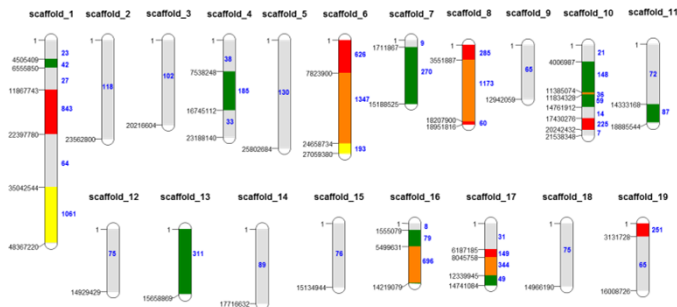
Résultats

Construction de la puce de génotypage 12K beads

Critères liés aux objectifs de génotypage

- Des régions candidates ciblées (QTL, gènes candidats) : 9169 SNP

- Une représentation homogène du génome hors régions ciblées : 1SNP/200kb – 1162 SNP



Critères techniques

- MAF ≥ 0.15 (15%) sur les 47 individus à faible couverture
- Validé sur au moins 1 des 4 individus à couverture $>20X$
- Élimination des SNP dans régions répétées ou dupliquées
- Pas de SNP/indel dans les 2 x 60 bases flanquantes
- Score Illumina $>0,8$ repetitive regions

Témoins au génotypage

51 individus reséquencés
1 témoin par plaque de 96
Témoins : 9 (parents + descendants)

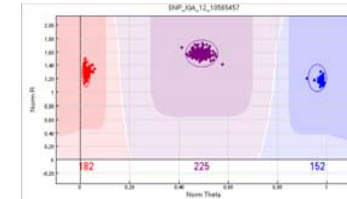
Résultats

Génotypage acquis avec la puce Illumina Infinium 12K beads

Illumina-Genome studio Software : Construction du/des cluster files

Elimination des échantillons avec un call rate < 0.95

Correction et classification du clustering - **nomenclature EPGV**



AUX1

Type	Nombre échantillons	Nombre d'individus
Interplaque	12 + 2	1
Génotypes reséquencés	75	51
Descendants	50	50
Parents	3	3
Individus	1010	1008
Individus éliminés		46
Total	1152	962

SNP commandés	10331
SNP "sur la puce"	9127 (88%)
AUX 1	7793 (85%)
AUX 2	159 (1,7%)
AUX 3	337 (3,6%)
AUX 4	453 (4,9%)
AUX 6	278 (3%)
AUX 7	101 (1,1%)
AUX 8	6 (0,06%)

Résultats

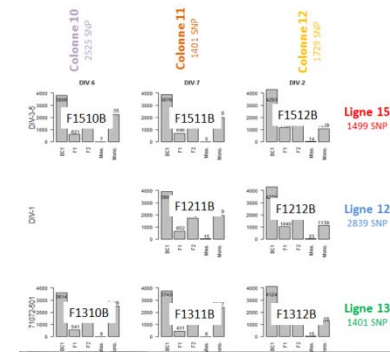
Génotypage acquis avec la puce Illumina Infinium 12K beads

Reproductibilité: témoin plaque (100%- n=12); Div-7 (>99%)

Ségrégation:

608 erreurs/ 411 877 allelic transmissions = 0,15%

Écart à une ségrégation mendélienne (1,65%)



Concordance Sanger-SNP / génotypage:

96%- 99% sur 259 SNP x 10 individus

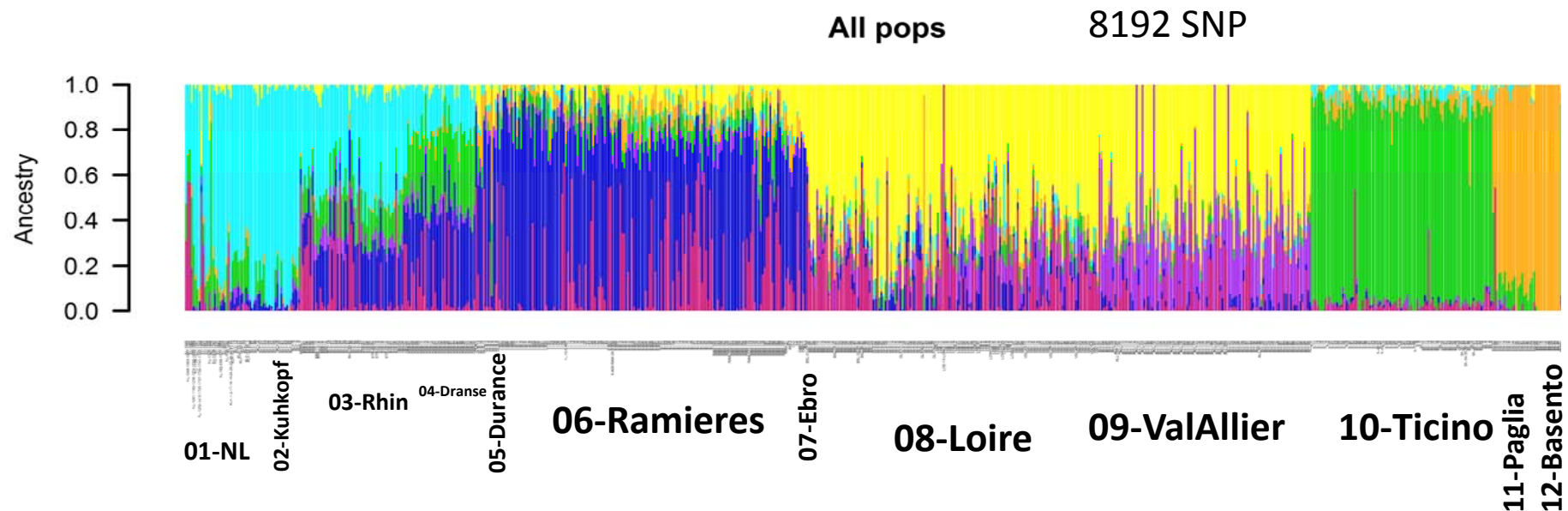
Concordance NGS-SNP / génotypage:

62% - 85% sur 7345 SNP x 16 individus

Erreurs correspondent à SNP_NGS homozygotes et SNP-génotypage hétérozygotes

Résultats

Structuration géographique de la diversité génétique au sein de la population d'association



- ✓ Une structuration moléculaire cohérente avec l'organisation géographique
- ✓ Des populations présentant un niveau d'admixture élevé surtout dans le Nord et l'Ouest
- ✓ Des stucturations proches avec 600, 2000 et 8192 SNPs

Conclusions-Perspectives (1)

1- NGS : Un moyen performant d'identifier des SNP

si une séquence de référence + $X > 10^{-12}$

Des informations à stocker dans des bases de données GniIS

2- NGS : études de variants structuraux en cours

Prédiction génomique, prédiction de l'hétérosis

3- Un assemblage *de novo* pour une référence *P.nigra*

Conclusions-Perspectives (2)

4- Puces de génotypage : des outils à construire à façon

qualité des ADN , coût puce 12K beads: 114 K€/ Point: 0.014 €

Suivi des flux de gènes

Suivi de la diversité dans les unités conservatoires

5- Génotypage haut-débit: des besoins encore plus importants

Génétique d'association avec une approche tout génome

Sélection génomique, mesures d'apparentement

Puce HD multi-espèces vs GBS ?

Puce HD multiespèces vs reséquençage à faible couverture?

Valorisation RNA-Seq transcriptome complet sur 240 génotypes

