

## Colloque EPGV 2016

# Génotypage SNP haut débit : *retour d'expériences sur puces AXIOM® Affymetrix 50k et 600k chez le tournesol*

Nicolas POUILLY - INRA Toulouse

# Equipe Génomique & Génétique du Tournesol

- **Tolérance à la sécheresse**  
Identification du contrôle génétique  
Approches systèmes (réseaux métaboliques, écophysioles)
- **Résistance aux pathogènes:** *Orobanche cumana*; *Plasmopara Halstedii*  
Identification du contrôle génétique (QTL, association)  
Analyse fonctionnelle de l'interaction (RNASeq, clonage positionnel)
- **Ressources génétiques:** Centre de Ressources Biologiques Tournesol  
Développement de nouvelles populations (EMS, MAGIC...)
- **Ressources génomiques:** Séquençage génome *H. annuus*  
Séquençage génome *O. cumana*
- **Développements d'outils de phénotypage:** plateforme HELIAPHEN



**SUNRISE**  
UNE CULTURE POUR LE FUTUR

**Objectif** : améliorer la production d'huile issue de la culture d'hybrides de tournesol en condition de disponibilité réduite en eau.

17 partenaires

Durée: 7 ans et 4 mois

Budget: 7 millions d'€ (subvention d'état)

10 work packages (WP)

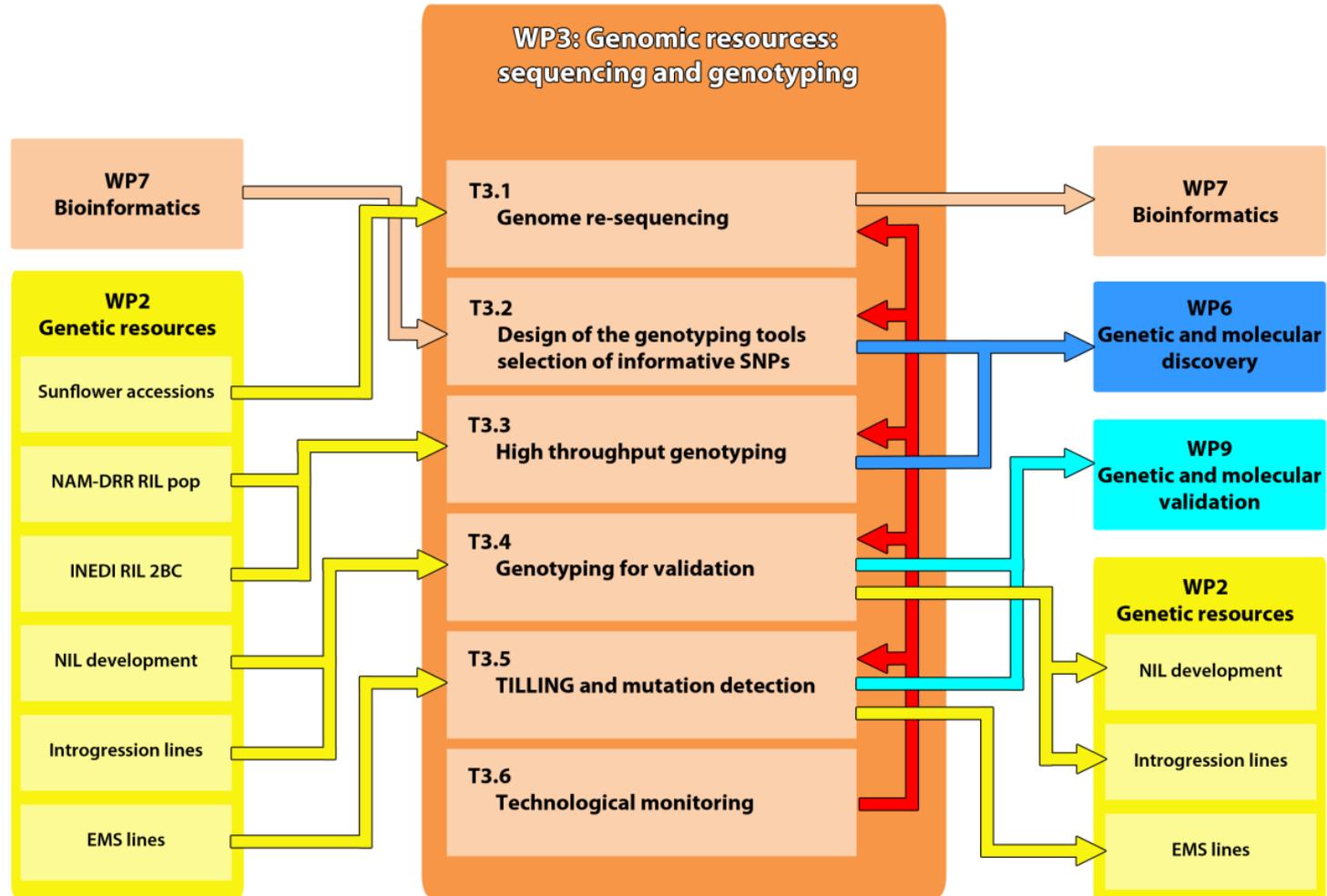
Coordinateur : [Nicolas.Langlade@toulouse.inra.fr](mailto:Nicolas.Langlade@toulouse.inra.fr)





**SUNRISE**

UNE CULTURE POUR LE FUTUR



# Dispositifs génétiques

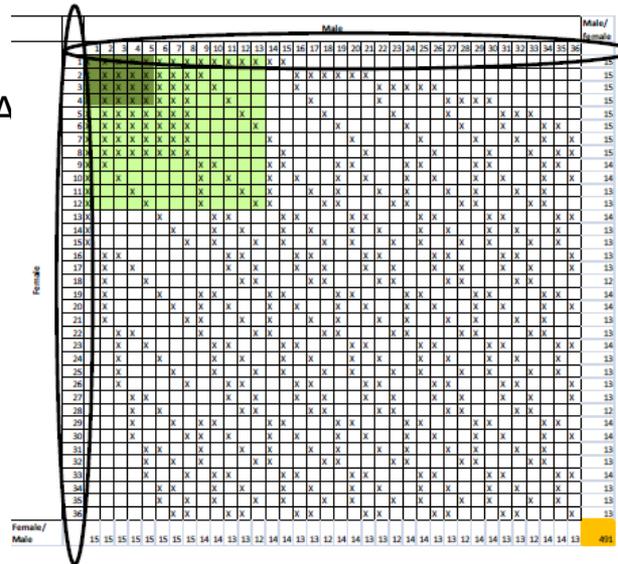
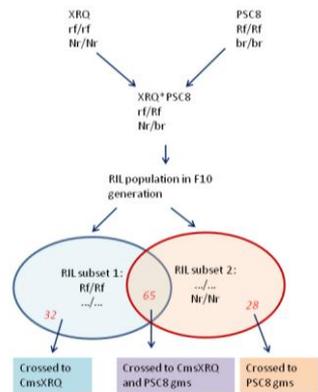
✓ **GWAS**

491 hybrides testés en conformité  
3 lignées élités / partenaire + 57 lignées INRA

✓ **NAM-DRR (18x32RILS=576RILs)**

cross	Parental lines								
	A	B	C	D	E	F	G	H	I
1	1	1	0	0	0	0	0	0	0
2	1	0	1	0	0	0	0	0	0
3	0	1	1	0	0	0	0	0	0
4	0	1	0	1	0	0	0	0	0
5	0	0	1	1	0	0	0	0	0
6	0	0	1	0	1	0	0	0	0
7	0	0	0	1	1	0	0	0	0
8	0	0	0	1	0	1	0	0	0
9	0	0	0	0	1	1	0	0	0
10	0	0	0	0	1	0	1	0	0
11	0	0	0	0	0	1	1	0	0
12	0	0	0	0	0	1	0	1	0
13	0	0	0	0	0	0	1	1	0
14	0	0	0	0	0	0	1	0	1
15	0	0	0	0	0	0	0	1	1
16	1	0	0	0	0	0	0	1	0
17	0	1	0	0	0	0	0	0	1
18	1	0	0	0	0	0	0	0	1

✓ **INEDI (125RILs 2BC)**



- Identifier des sites polymorphes sur les 72 parents pour prédire le génotype de chaque hybride du dispositif GWA.
- Sélectionner des SNPs pour la production d'une puce AXIOM 600k pour le génotypage des dispositifs NAM-DRR et INEDI.

# AXIOM 600k SNPs: sélection des SNPs

- **Re-séquençage de 72 lignées *H. annuus***

Technologie Hi-Seq: PE 2x100nt, profondeur 10X / lignée

Cartographie sur assemblage du génome (2014): lignée XRQ  
199944 scaffolds  
N50: 24.4kb  
Taille totale: 1.58Gb

- ✓ Filtre des SNPs:

- au moins 10 lignées *H. annuus* avec l'allèle de référence XRQ
- au moins 10 lignées *H. annuus* avec l'allèle variant
- au moins 1 lignée *H. annuus* avec allèles de référence ou variant présent sur les 2 brins séquencés
- SNPs bi-allélique uniquement

- **6 348 868 SNPs pour 14129 scaffolds génomiques**

- ✓ Sélection des SNPs:

- sélection de plusieurs SNPs / scaffold, pour avoir au moins un SNP polymorphe pour chacun des 18 croisement du NAM-DRR
- Pas de polymorphisme sur la séquence contexte de 20pb autour du SNP

- **586 985 SNPs pour 14 129 scaffolds sur puce AXIOM 600k**

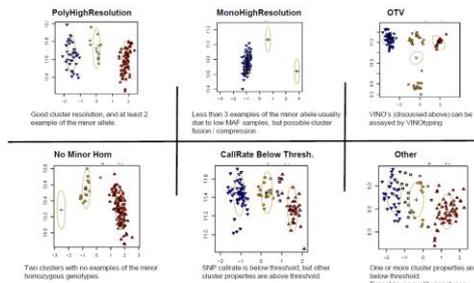
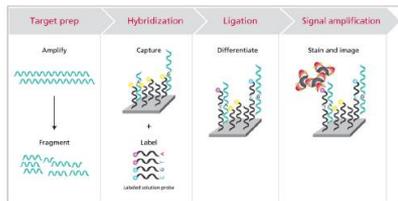
# AXIOM 50k SNPs: sélection des SNPs

- **AXIOM 600kSNPs** sur un panel de **457 lignées** (*source: SYNGENTA*)
  - 146142 monomorphes
  - **439277 polymorphes (75%)**
  - 1566 avec données manquantes > 60%
- ✓ échantillonnage sur 439277 SNPs / 14021 scaffolds génomiques
- ✓ sélection avec données manquantes <3%; MAF >10%: **163707 SNPs**
- ✓ sélection de 2 SNPs par scaffold génomique, pour maximiser leur spécificité de discrimination
- **AXIOM 50k: 46526SNPs + 2925 SNPs** (WP3 PIA SUNRISE)

# Génotypage AXIOM



**GENTYANE**  
PLATE-FORME de  
**GENOTYPAGE et SÉQUENÇAGE en AUVERGNE**



	50k	600k
<b>Nombre d'échantillons / puce</b>	384	96
<b>Nombre de SNPs</b>	49451	586985
<b>Nombre de scaffolds géno.</b>	14021	14129
<b>Prix / échantillon</b>	42 €	208 €
	INRA, Syngenta Fr, RAGT2n, Caussade Sem. Maisadour Sem.	

# Génotypage AXIOM 600k

- **Cartographie Génétique**

- ✓ Validation des marqueurs: - parents homozygotes et polymorphes
  - hybride F1 hétérozygotes
  - fréquence allélique : 40-60%
  - données manquantes <5% et hétérozygotes (AB) <5%
- ✓ Carte génétique: Cartagène, avec fusion des marqueurs colocalisés

- Population INEDI (PIA SUNRISE): 228 RILs + 2 XRQ + 2 PSC8 +2 INEDI (F1)  
Matrice retenue: 55 950 SNPs  
Carte génétique: 2240 bins génétiques/1489,4 cM
- Population HA89\*LR1 (Louarn, 2016): 101 RILs + 3 HA89 + 3 LR1  
Matrice retenue: 21201 SNPs  
Carte génétique: 951 bins génétiques/1795,8 cM

# Génotypage AXIOM 600k

- Approche de type SAM pour le développement de lignées d'introggression (IL) (PIA SUNRISE)

**Plateforme de diversité allélique: 5 populations d'IL XRQ \* H. annuus sauvages (100IL/pop)**

Screening par set de 384SNPs lors de la création des populations

*BC1: génotypage 400 plantes/pop génotypée, sélection de 100 plantes*

*BC2: génotypage 400 plantes/pop génotypée, sélection de 100 plantes*

*BC2S1 : génotypage 200 plantes/pop génotypée, sélection de 100 plantes*

- ✓ Génotypage des parents sauvages des ILs et XRQ
- ✓ Identification des SNPs polymorphes chaque population pour SAM

ACC_CO D	FOUR_C OD	FOUR_A LT	nb wt parent	SNPs_mappés_ca rte_454
649	ANN 1184	PI468468	2	27306
654	ANN 1430	PI468576	3	10718
826	ANN 1510	PI468620	2	34580
831	ANN 1064	PI468439	8	7694
833	ANN 1438	PI468581	2	33277

# Génotypage AXIOM 600k

- **Identification de marqueurs moléculaires liées à des régions d'intérêt** (*coll. CSIC, Cordoba*)
  - ✓ Stratégie de bulk segregant analysis (BSA) pour:- résistance Orobanche  
- QTL Teneur en tocophérols
  - ✓ BSA réalisée avec 2 lignées parentales + 2 bulks (Resistant/Sensible; High/Low)

Marqueurs AXIOM	Orobanche	Tocophérol
Monomorphes (Parents)	284 729	225 384
<b>Polymorphes (Parents)</b>	<b>67 374</b>	<b>50 034</b>
Discrimants au moins 1 bulk AB	53 308	39 269
Discrimants les 2 bulk AB	39 319	26 211
Liés au caractère	563	939
Positionnés sur carte génétique (2014)	272	614
<b>Positionnés sur région du LG</b>	<b>260 (95,5%)</b>	<b>560 (91%)</b>

# Génotypage AXIOM 50k

- **Ressources Génétiques du genre Helianthus (HELIADIV, PROMOSOL)**

**RG de type population: 6 POP \* 80ind**

**RG de type lignées: 534 lignées**

**20729 SNPS PolyHighRes**

- ✓ **Concordances intra population: identification d'une POP non conforme**

Genotype_id2 for Reference	Genotype_id2 for Base	%Concordance
HDV_POP077_LIPM3492-B04	2603	99,85
HDV_POP077_LIPM3492-E11	2603	99,83
HDV_POP077_LIPM3492-A12	2603	99,78
HDV_POP077_LIPM3492-A01	2603	99,75
HDV_POP077_LIPM3492-E03	2603	99,64
HDV_POP077_LIPM3492-A08	2603	99,61
HDV_POP077_LIPM3492-B07	2603	99,49
HDV_POP077_LIPM3492-E10	2603	99,11

# Génotypage AXIOM 50k

- **Ressources Génétiques du genre Helianthus (HELIADIV, PROMOSOL)**

**RG de type population: 6 POP \* 80ind**

**RG de type lignées: 534 lignées**

**20729 SNPS PolyHighRes**

- ✓ **Concordances intra population: identification d'une lot d'une POP non conforme**

- ✓ **Concordances entre lignées**

Genotype_id2 for Reference	n_samples	%Concordance	comment
XRQ	3	99,74	
AA	2	99,81	
ADV	2	98,85	
DI	2	99,47	
IR	2	99,44	
PW3RM	2	99,56	
PM17	2	99,59	
OQP8	2	98,17	
RHA438	2	96,45	2 lots différents
ZYQ	2	96,44	2 lots différents

# Génotypage AXIOM 50k

- Ressources Génétiques du genre Helianthus (HELIADIV, PROMOSOL)

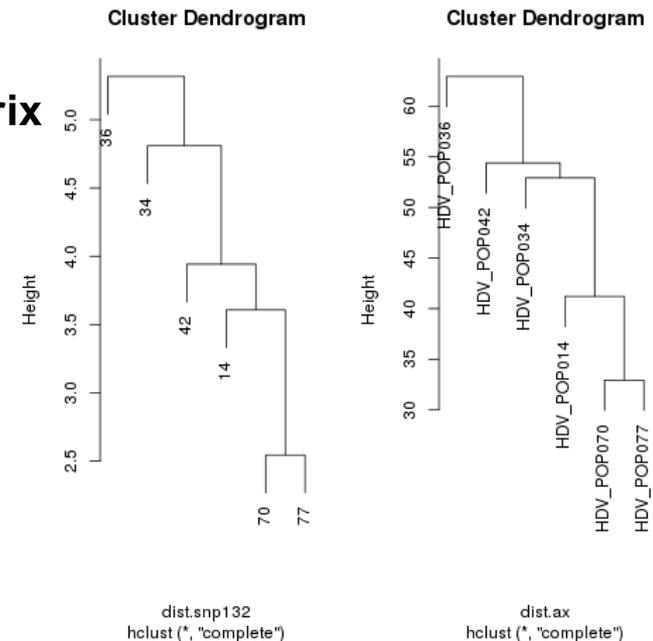
RG de type population: 6 POP \* 80ind

RG de type lignées: 534 lignées

20729 SNPS PolyHighRes

- ✓ Concordances intra population: identification d'une lot d'une POP non conforme
- ✓ Concordances entre lignées
- ✓ Comparaison BeadXpress@Illumina vs AXIOM@Affymetrix

- 132 SNPs BXP                    11 échantillons / POP
- 18 746 SNPs AXIOM        80 échantillons / POP
- matrices de fréquences alléliques



# Génotypage AXIOM 50k vs 600K

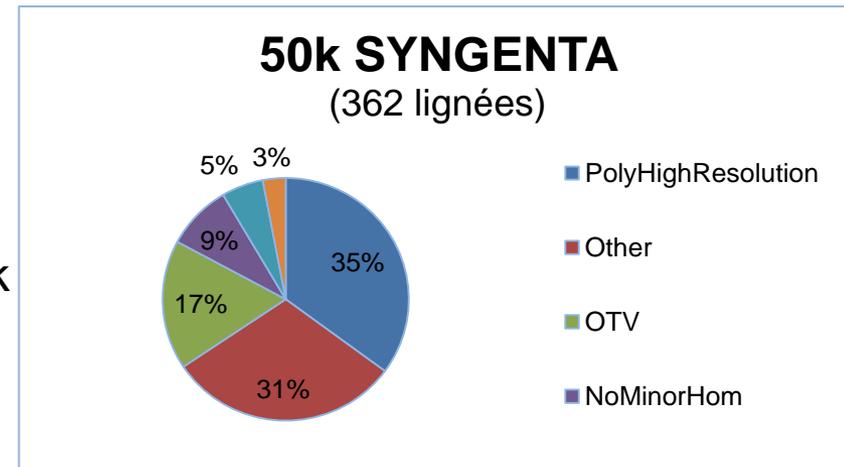
- **SYNGENTA:**

Axiom 600k 485795 SNPs

Axiom 50k 46941 SNPs (commun avec 600k)

29 lignées communes sur puces 600k et sur 50k

- ✓ Comparaison des données entre les 2 types de puces
- ✓ Recherche des marqueurs non concordants de la puce 50k pour lesquels il manquait un des deux allèles homozygotes (ie marqueur monomorphe ou problème de clustering): 15916 SNPs
- ✓ Si ce nombre est du à un problème de clustering, on devrait en trouver en commun avec les marqueurs problématiques identifiés par comparaison des données de la 600k et de la 50k.



Nombre minimal d'individus avec une donnée non concordante 50k/600k	Nombre de marqueurs non-concordants (sur les 46941 en commun)	Marqueurs en commun
1	19854	10184 (51%)
5	8873	6212 (70%)
10	5042	3719 (74%)
15	2793	2298 (82%)
20	1724	1573 (91%)
25	995	959 (96%)
29	328	328 (100%)

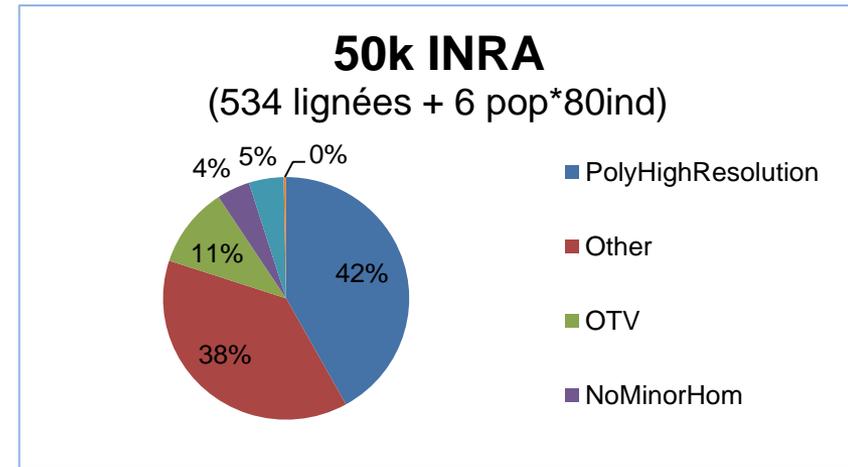
# Génotypage AXIOM 50k vs 600K

- **INRA:**

Axiom 600k 485795 SNPs

Axiom 50k 20729 PolyHigh (commun avec 600k)

2 lignées de référence: XRQ – PSC8



Lignées	AA-BB	AA-AB-BB	Marqueurs communs
XRQ	5 (0,03%)	550 (3,60%)	265
PSC8	9 (0,05%)	688 (2,88%)	

➤ en cours:

- hybridation d'hybrides pour améliorer le clustering
- question à Affymetrix sur la définition des library files de la puce Axiom 50k

# Conclusions

- Investissement initial nécessaire
- Facilité pour la réalisation de différents projets de génotypage
- Rester parcimonieux sur les résultats bruts
- Support Affymetrix
- *Perspectives:*
  - Dispositif 576 NAM-DRR
  - Core-collection 96 *H. annuus* pour étude association Orobanche
  - Cartographie nouvelles résistance Mildiou par BSA

 **Génétique & Génomique du Tournesol**

Stéphane Muñoz  
Marie-Claude Boniface  
Nicolas Langlade  
Nicolas Blanchet  
Anne-Sophie Lubrano  
Romain Dinis  
Brigitte Mangin  
Alexandra Duhnen  
Fanny Bonnafous  
Florie Gosseau  
Prune Pegot-Espagnet  
Johann Louarn  
Pauline Duriez  
Xavier Grand  
Gwenola Marage  
Marion Larroque  
Laurence Godiard  
Yann Pecrix  
Charlotte Penouilh  
Olivier Catrice

 **Bioinformatique**

Jérôme Gouzy  
Sébastien Carrère  
Ludovic Legrand  
Erika Sallet



**GENTYANE**  
PLATE-FORME de  
GENOTYPAGE et SÉQUENÇAGE en AUVERGNE

Charles Poncet  
Anthony Théron



**SUNRISE**  
UNE CULTURE POUR LE FUTUR

**P.R.O.M.O.S.O.L.**  
ASSOCIATION  
POUR LA PROMOTION  
DE LA SÉLECTION  
DES PLANTES OLÉAGINEUSES