



HAL
open science

Analyse de diversité d'un panel de pré-breeding de blé tendre par une approche transcriptomique

Hélène Rimbert, Frédéric Choulet, Odile Argillier, Jérôme Auzanneau, Mark Davey, Philippe Dufour, Sylvie Dutriez, Pascal Giraudeau, Ellen Goudemand-Dugue, Gemma Molero, et al.

► To cite this version:

Hélène Rimbert, Frédéric Choulet, Odile Argillier, Jérôme Auzanneau, Mark Davey, et al.. Analyse de diversité d'un panel de pré-breeding de blé tendre par une approche transcriptomique. 9^{ème} rencontres du FSOV -2024, Apr 2024, Paris, France. hal-04712172

HAL Id: hal-04712172

<https://hal.inrae.fr/hal-04712172v1>

Submitted on 27 Sep 2024

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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Hélène RIMBERT¹, Frédéric CHOLET¹, Odile ARGILLIER², Jérôme AUZANNEAU³, Mark DAVEY⁴, Philippe DUFOUR⁵, Sylvie DUTRIEZ⁶, Pascal GIRAudeau⁷, Ellen GOUEMAND-DUGUE⁸, Gemma MOLERO⁹, Mickaël THROUDE¹⁰, Hervé DUBORJAL¹⁰, Adeline CLEMENTI¹⁰, David GRIMBICHLER¹¹, Etienne PAUX¹², Sophie BOUCHET*¹

1. INRAE Université Clermont-Auvergne, UMR 1095, GDEC, 5 chemin de Beaulieu 63100 Clermont-Ferrand, France
2. Syngenta France SA, 2 avenue Gustave Eiffel, F-28000 Chartres, France
3. AGRI-Obtentions, Chemin de la Petite Minière, 78280 Guyancourt, France
4. BASF Innovation Center Gent, Technologiepark-Zwijnaarde 101, 9052 Gent, Belgique
5. RAGT, Rue Emile Singla, BP 3331 12033 Rodez Cedex 9, France
6. Lidea Seeds, avenue Gaston Pheobus, 64230 Lescar, France
7. Secobra Recherches, Centre de Bois-Henry, 78580 Maule, France
8. FLORIMOND DESPREZ VEUVE & FILS, 59242 Cappelle-en-Pévèle, France
9. KWS MOMONT SAS, 7 Rue de Martinal, 59246 Mons-en-Pévèle, France
10. Limagrain Europe, Centre de recherche de Chappes, 63720 Chappes, France
11. UCA, Plateforme AuBi & Mésocentre Clermont-Auvergne, 63000 Clermont-Ferrand, France
12. VetAgro Sup, 89 Avenue de l'Europe, CS 82212, 63370 Lempdes, France



Introduction

Dans le cadre du projet DéBAT, nous avons produit un catalogue de l'ensemble des gènes exprimés dans les lignées d'un panel de 450 lignées représentatives de la diversité mondiale par la méthode RNA-seq. Douze lignées représentatives ont été séquencées en forte profondeur sur 3 tissus (tige, épi et feuille). Des niveaux d'expression différentiels ont été identifiés. Le reste du panel a été séquencé sur le mélange des 3 tissus. Ceci nous a permis d'identifier la présence absence (PAV) d'isoformes de pour 60K gènes annotés High Confidence sur la séquence de référence Chinese Spring et 83K gènes absents de cette séquence. Leur fonction a été prédite. Des analyses d'association ont été conduites sur les PAV et les SNP obtenus avec les phénotypes évalués dans le cadre du projet compagnon Ex-IGE.

Echantillonnage des lignées (Figure 1)-

Volet 1 - 12 lignées représentatives du panel BWP3 du projet Breedwheat ont été sélectionnées (Figure 1B) pour un séquençage RNA-seq forte profondeur (2 x 115 millions de reads, NovaSeq6000). **Volet 2** - Le transcriptome des 450 lignées du BWP3 (Figure 1A) a été séquencé en plus faible couverture (2x25 millions de reads).

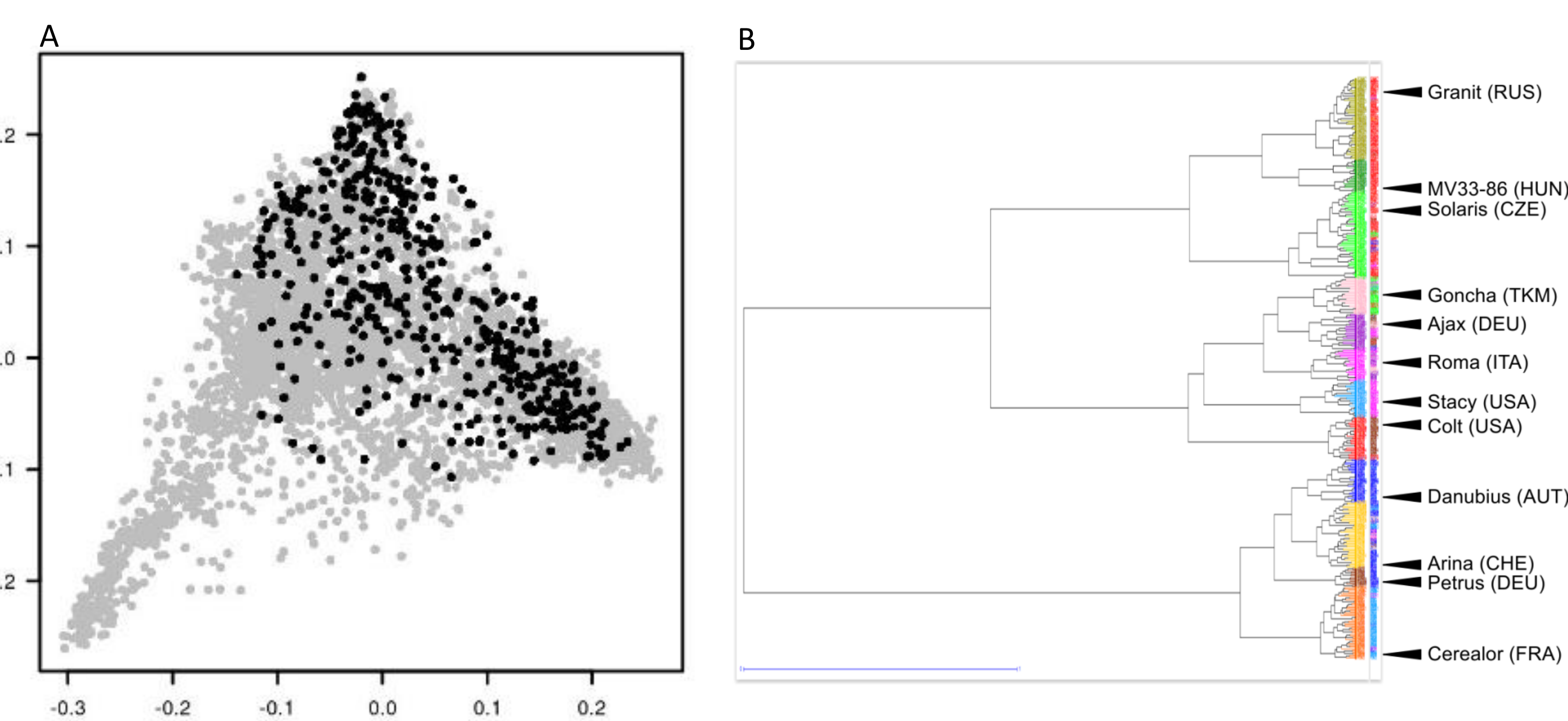


Figure 1 - (A) PCOA de 4500 lignées du CRB, représentatives de la diversité mondiale. Les points noirs correspondent aux 450 lignées du projet, adaptées à une évaluation en France; (B) Dendrogramme de Ward des 450 lignées. D'après Balfourier et al. (Science Adv 2019)

Pipeline d'analyses (Figure 2)-

Volet 1 - Le transcriptome des 12 lignées de référence a été séquencé dans trois tissus (tige, feuille, épi). Un pan-transcriptome de référence avec 400K reads unique a été construit. **Volet 2** - Le transcriptome des 450 lignées du BWP3 a été séquencé en mélange. Les reads ont été mappés sur le pan-transcriptome de référence construit dans le volet 1.

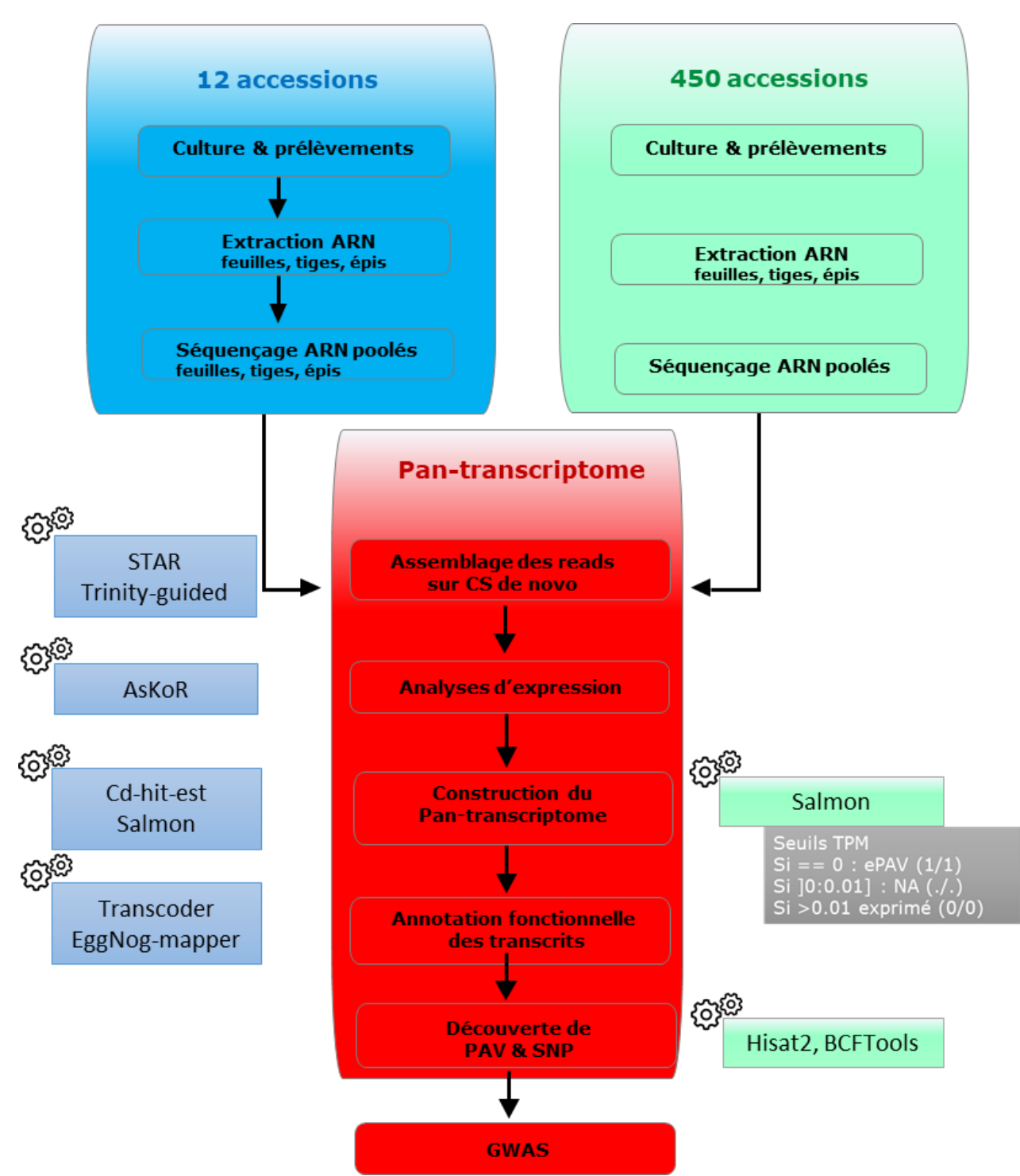


Figure 2: Pipeline d'analyses

Analyse d'expression (Figure 3)-

Volet 1 - Les gènes sont différentiellement exprimés selon les tissus.

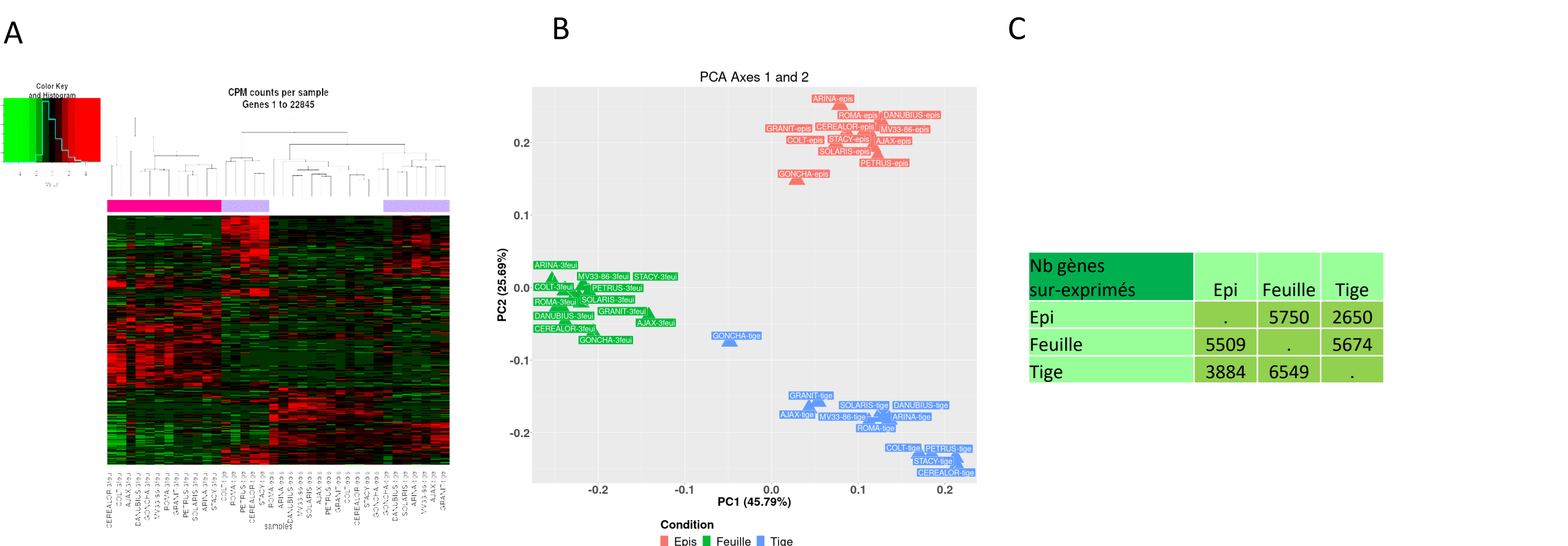


Figure 3 - Niveaux d'expression des gènes dans 12 lignées de référence et 3 tissus; (A) Dendrogramme; (B) ACP; (C) Nombre de gènes sur-exprimés dans les compartiments situés en ligne

Mapping des transcrits -

Volet 1 - Un pan-transcriptome de 400 288 transcrits a été construit à partir des 12 lignées de référence.

En moyenne, 300 millions de reads ont été séquencés par lignée de référence représentant 400K transcrits uniques. Ils correspondent à 60K gènes annotés chez CS (HC ou LC), soit 51% des gènes annotés (Figure 4A) et 84K nouvelles protéines dont nous avons prédit la fonction. Un nombre de reads équivalent est exprimé dans tous les tissus.

Volet 2 - La plupart des 400K pan-transcrits de référence (99%) sont exprimés dans au moins une des 450 lignées (Figure 3B2). En moyenne, 186k transcrits sont exprimés dans chaque lignée (Figure 4B1). Au total, 60K gènes sont exprimés en moyenne par lignée (Figure 4B2).

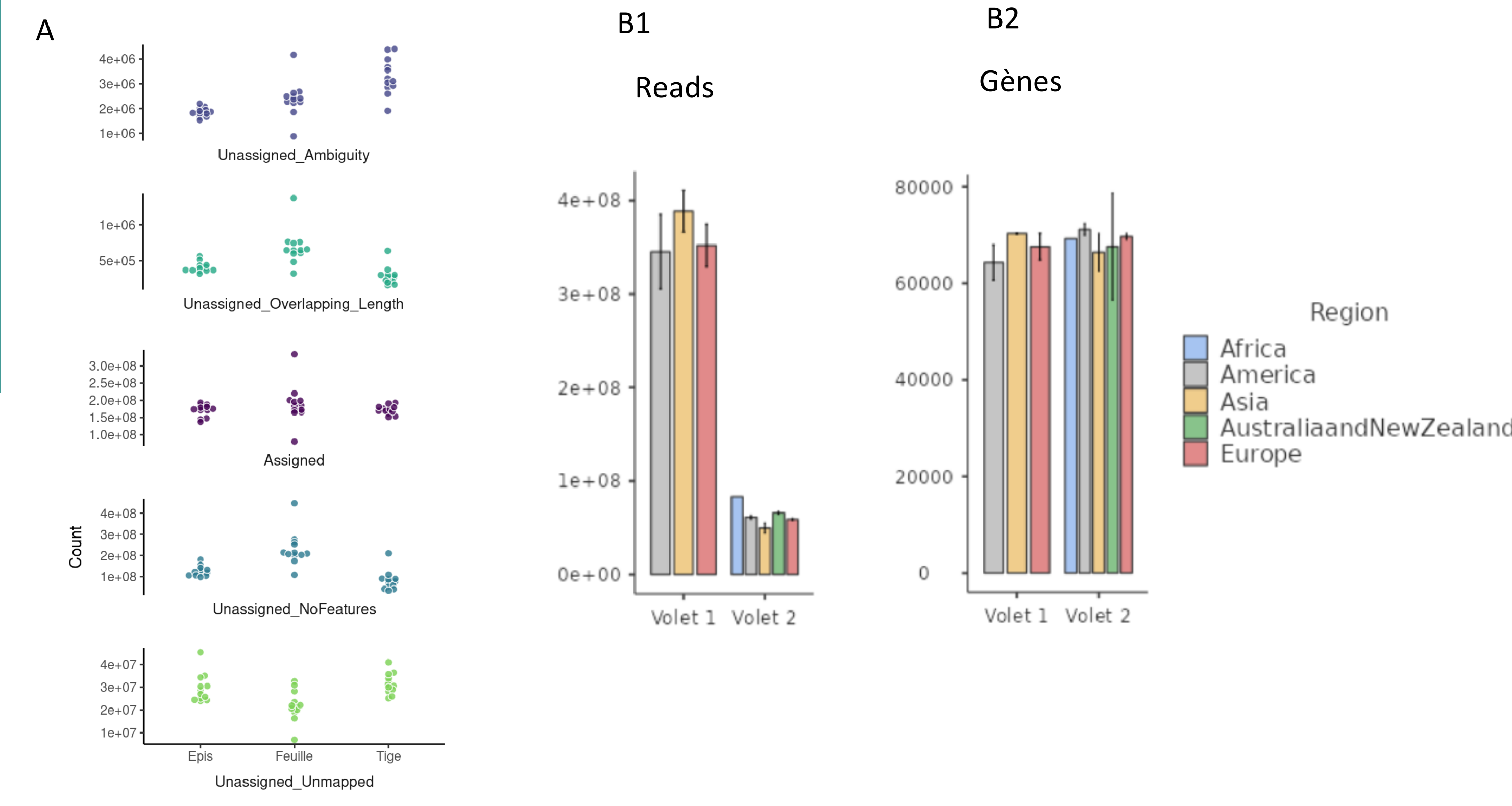


Figure 4 - (A) Assignment des reads RNA-seq "Featurecount" sur le génome de référence IWGSC - Chinese spring. Unassigned_Ambiguity: les 2 paires de reads ne mappent pas sur le même chromosome; Unassigned_Overlapping_Length: les deux paires de reads se chevauchent (minimum de 100 nucléotides de séparation autorisé); Assigned: mappent sur des régions de Chinese-Spring annotées; Unassigned_NoFeatures: mappent sur des régions de Chinese-Spring non annotées; Unassigned_Unmapped: ne mappent pas sur Chinese-Spring; (B) Nombre de reads et de gènes présents par zone géographique

Pan-transcriptome (Figure 5):-

Chaque transcrit a été attribué aux compartiments "core" (plus de 90% des lignées partagent le transcript), "shell" (entre 10 et 90%) ou "cloud" (moins de 10%). Au total, 71% des transcrits sont présents dans le "shell", 17% dans le "core" et 11% dans le "cloud". En moyenne, un transcrit est présent chez la moitié des lignées du panel.

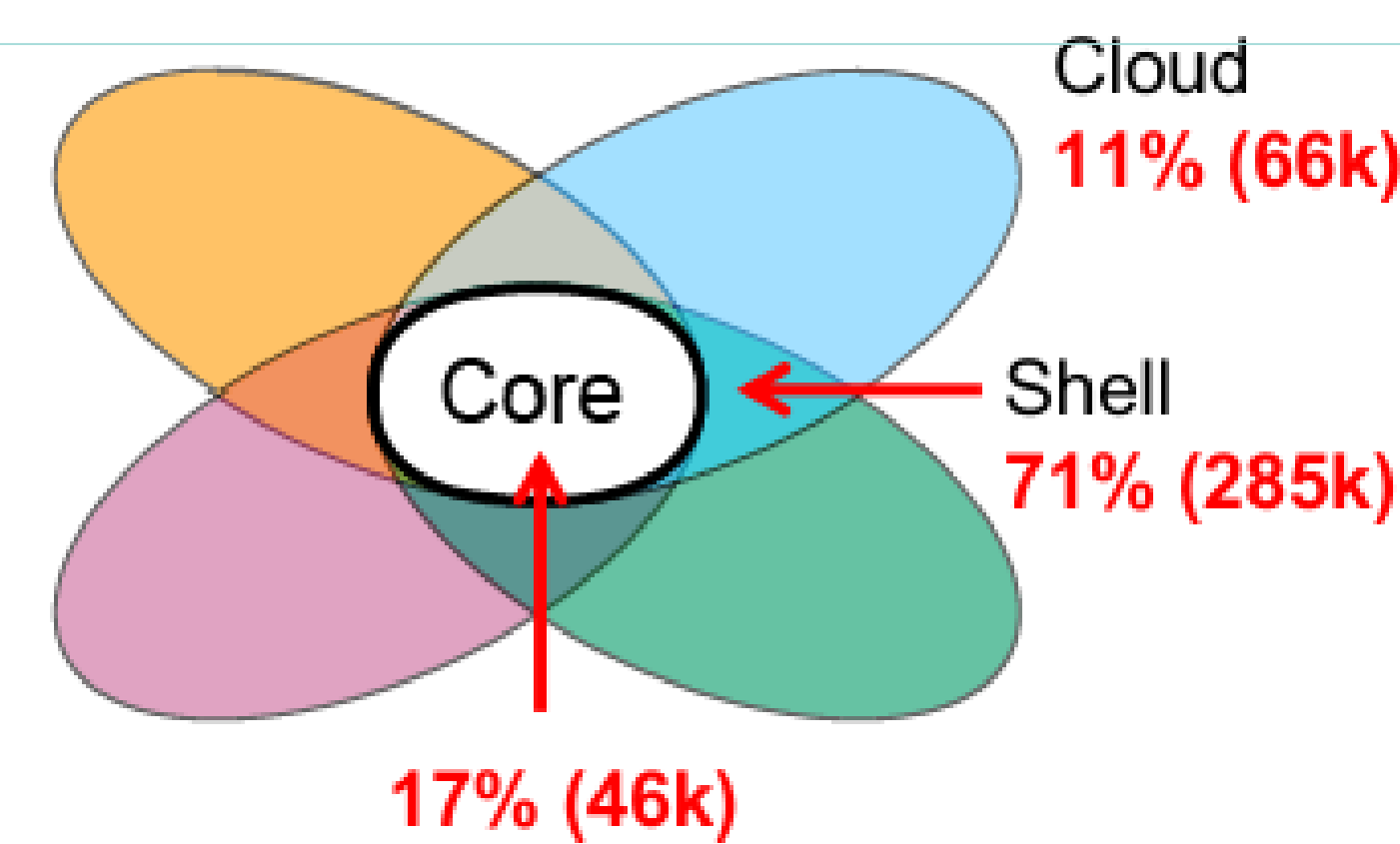


Figure 5 - Distribution du pan-transcriptome

Analyse fonctionnelle - La construction du pan-transcriptome a permis l'identification de **84714 nouvelles protéines** pour lesquelles une fonction a été prédite (Transdecoder, eggNOG-mapper). L'enrichissement en GO de ces protéines (R-topGO) a montré une prévalence pour les processus adaptatifs (stress biotiques, abiotiques, lumière, température, stress osmotique ...).

ePAVs et SNP calling-

Nous avons utilisé un seuil de 0,01 TPM pour déclarer un transcrit présent dans une lignée (Figure 6). Les transcrits avec un TPM compris entre 0 et 0,01 ont été déclarés en données manquantes. Après filtre sur données manquantes (<0,1) nous avons gardé 362 899 marqueurs ePAV. Un SNP calling a été fait parmi les 400K transcrits. Après filtre sur données manquantes, nous avons gardé 83K SNP. Les Minimum Allele Frequency (MAF) sont équilibrées pour les ePAV. On observe un excès d'allèles rares pour les SNP (Figure 7). En moyenne, les SNPs comptent 8% de données manquantes, et les ePAV présentent très peu de données manquantes (au maximum 2% par marqueur). Les SNP issus des ePAV discriminent les mêmes groupes génétiques que les SNP Breedwheat (Figure 8). Par contre les ePAV discriminent des groupes différents.

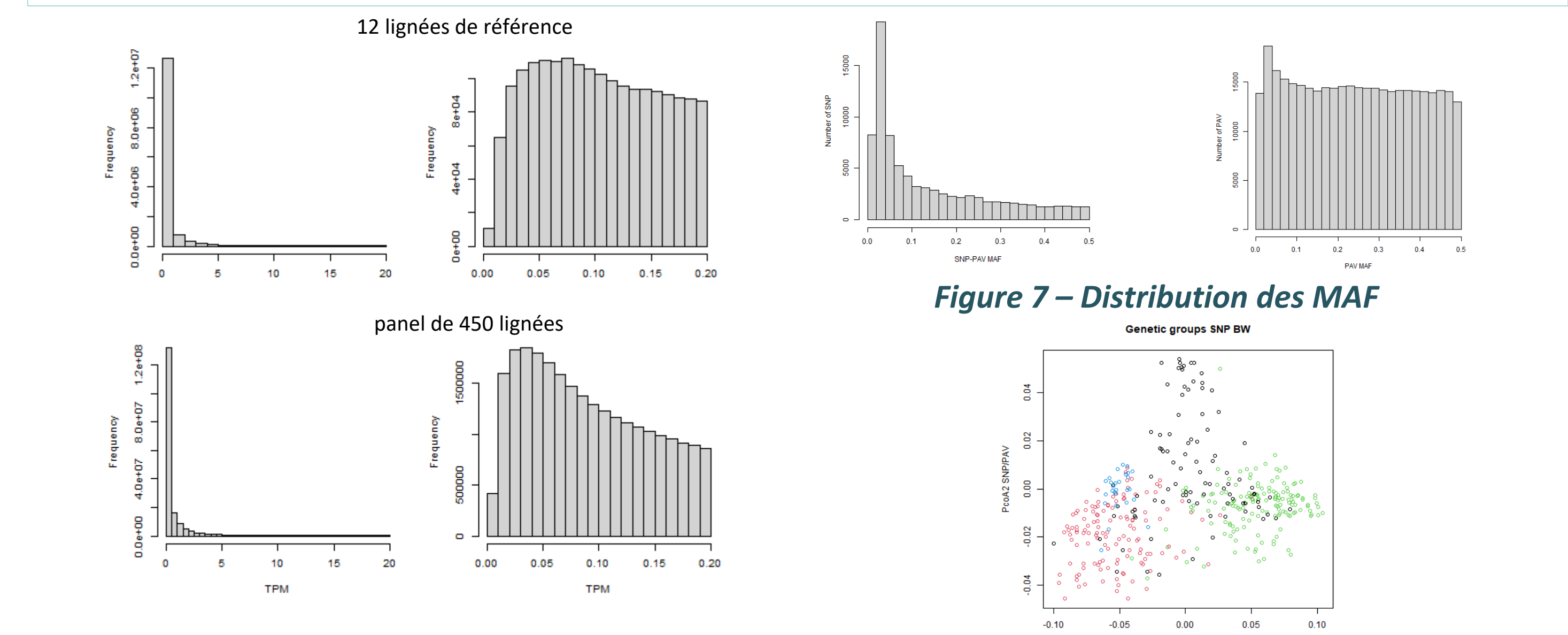


Figure 6 - Distribution des TPM; Figure 7 - Distribution des MAF; Figure 8 - PCOA construite avec les SNP issus de ePAV

Détection automatique d'introgessions -

Une analyse est en cours pour détecter automatiquement les introgessions. Par exemple l'introgession de seigle 1BS/1R est détectée chez Solaris avec une expression nulle des gènes de blé tendre sur le bras court (Figure 9).

Perspectives -

Des études d'associations ont été conduites avec les marqueurs produits dans ce projet et les phénotypes produits de le projet compagnon Ex-IGE. En éliminant les marqueurs SNP avec beaucoup de données manquantes, nous avons probablement éliminé tous les marqueurs avec trois allèles (absence et présence avec SNP). Il serait intéressant d'exploiter toutes les données Breedwheat et DéBAT pour valider notre méthode de détection sur des introgessions connues. Nous regarderons également si parmi les reads éliminés ou "unmapped", certains correspondent à des reads d'espèces apparentées (ventricosa, thimopyrum, dicoccoides, monococcum, speltoides, seigle...).

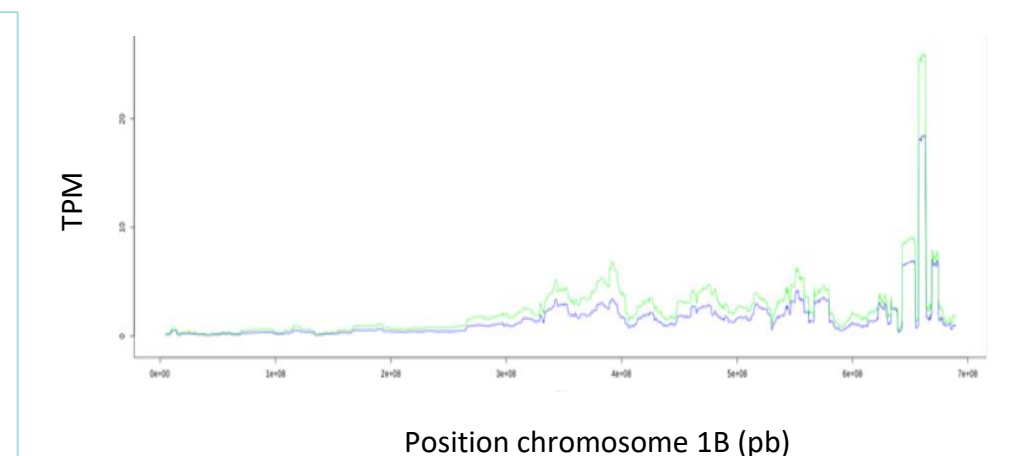


Figure 9 - Profil d'expression des gènes du chromosome 1B chez Solaris