IWGSC RefSeq v2.0 Annotation

Frédéric Choulet
GDEC, INRAE, UCA, Clermont-Ferrand, France
Task force:
GDEC (Rimbert, Leroy, Choulet)
PGSB (Twardziok, Lang, Spannagl)
Earlham I. (Swarbreck, Venturini)

RefSeq v1.0
(2018)

Anno v1.0

Anno v1.1

RefSeq v2.0
(2019)

in progress...

Anno v1.2

2461 curated genes
(contributions received in 2019)

Guideline for Anno_v2.0:
- Transfer annotated genes from v1.2 (as much as possible...)
- Avoid *denovo* annotation → limited to new regions
- Keep gene IDs !!!
Anno_v2.0

Unable to calculate new positions of features directly on RefSeq v2.0
- **Genes** -> use homology-based approach
  - 2.9 M features (~300k transcripts HC+LC)
- **TEs** -> *denovo*

- **Phase 1:** estimate the level of ambiguity using homology-based transfer of annotation
  - 15% (**45k**) mRNAs cannot be mapped perfectly

- **Phase 2:** develop a dedicated approach to maximize the number of genes perfectly mapped

- **Phase 3:** resolve non-perfectly mapped genes?!

- **Phase 4:** annotate filled gaps (new sequences)
- Release v2.0 through URGI JBrowse

Release v2.0 through URGI JBrowse