

THE CULTIVATED SUNFLOWER PAN GENOME PROVIDES INSIGHTS ON THE WILD SOURCES OF INTROGRESSIONS AND THEIR ROLE IN BREEDING.

***Sariel HUBNER*¹, *Emanuel ZIGLER*², *Jennifer R. MANDEL*³, *Dirk SWANEVELDER*⁴, *Patrick VINCOURT*⁵, *Nicolas LANGLADE*⁶, *John M. BURKE*⁷, *Loren Henry RIESEBERG*¹**

¹ *University of British Columbia*

² *SAP*

³ *University of Memphis*

⁴ *Biotechnology Platform Agricultural Research Council*

⁵ *INRA - LIPM*

⁶ *INRA*

⁷ *University of Georgia*

sariel.hubner@botany.ubc.ca

ABSTRACT

Since domestication, cultivated sunflower has accumulated gene introgressions from its direct ancestor and closely related species. Wild relatives of sunflower are known for their enhanced tolerance for both biotic and abiotic stress and therefore are a promising genetic resource for breeding robust varieties. As the wild source used for crossing in breeding programs varies, different introgressed segments are obtained in different lines. The outcome of these introgressions is variation among lines in the genomic composition and presence/absence of specific genes. Therefore, some genes found in the cultivated gene pool are absent from the sunflower reference genome corresponding to a specific line, HA412. Identifying the overall repertoire of genes across lines, referred as the species pan genome, is of great interest for both evolutionary biologists and breeders. Here we present the cultivated sunflower pan genome based on the public association mapping population, which is comprised of 288 lines that capture most of the diversity in the cultivated gene pool. Sequences from each accession were aligned to the reference genome and badly mapped reads were extracted and used for *de-novo* assembly of sequences not found in the reference genome. Assembled sequences were further annotated to reach a unified non-redundant set of 21,081 sequences corresponding to the dispensable fraction of the cultivated sunflower pan genome. To identify the potential introgressions, low coverage whole genome sequences from 192 accessions of different wild *Helianthus* taxa were aligned to the dispensable portion of the pan genome, and 831 genes of wild parentage were identified. Thus, wild introgressions have contributed both new alleles and new genes to the cultivated sunflower gene pool. We further explore the wild origin of each introgressed gene and identify their underlying contribution to phenotypic variation in the cultivated sunflower gene pool.

Key Words : Pan genome, Introgressions