

Résumé

La vigne (*Vitis vinifera*) possède un métabolisme secondaire particulièrement riche donnant naissance à une large palette de molécules dont certaines sont impliquées dans les défenses contre les pathogènes et d'autres dans la grande diversité d'arômes qui fait la renommée des vins. L'analyse de la séquence de référence du génome de la vigne a permis de mettre en évidence une remarquable expansion de certaines familles de gènes liés au métabolisme secondaire par rapport aux autres plantes. Dans ce travail, j'ai étudié les familles de gènes codant pour les *cytochromes P450*, dont certains sont impliqués dans la production d'arômes, les gènes codant pour les stilbènes synthases (*STS*), les endo- β -1,3-glucanases et les gènes de résistance de type *NBS* impliqués dans les défenses de la vigne. Ma thèse vise à proposer des hypothèses expliquant l'organisation structurale de ces familles de gènes et ainsi à mieux comprendre pourquoi certaines familles présentent une amplification dans le génome de la vigne.

Des approches bioinformatiques ont été utilisées afin d'étudier ces différentes familles de gènes. Les gènes *cytochromes P450* et gènes R de type *NBS* ont tout d'abord été annotés de manière manuelle dans le génome de référence de la vigne. L'expression des gènes endo- β -1,3-glucanases, *STS* et *cytochromes P450* a été analysée en utilisant une approche transcriptomique à grande échelle. Pour ce faire, un outil a été développé durant cette thèse pour estimer le niveau d'expression des gènes à partir de données RNA-Seq disponibles dans les banques de données publiques. Parallèlement, des données de reséquençage d'ADN de 56 cépages et espèces de vigne ont été analysées, afin de déterminer les variations structurales de type CNV au sein des familles de gènes à domaine *NBS* et de gènes *STS*.

Ces différents travaux ont permis de montrer que l'amplification des familles de gènes étudiées n'est pas spécifique du génome de référence mais est retrouvée dans l'ensemble du genre *Vitis*, mais également de mettre en évidence des variations structurales au sein des différents génomes étudiés. L'analyse de la famille *STS* a montré que ces gènes sont organisés en blocs de duplication, et que les gènes plus conservés sont aussi les plus exprimés. Nous avons également montré que les gènes à domaine *NBS* sont organisés en cluster, dont certains sont particulièrement soumis à variation. Ces travaux contribuent à une meilleure connaissance de facteurs de défense efficaces et durables ainsi que des gènes impliqués dans la synthèse d'arômes dans la vigne. Ces connaissances pourront bénéficier aux programmes de création variétale mis en œuvre à l'INRA de Colmar.

Mots clés : vigne, familles de gènes, génomique comparative, transcriptomique, CNV

Abstract

Grapevine (*Vitis vinifera*) has a particularly rich secondary metabolism, giving rise to a wide range of molecules, some of which are involved in defences against pathogens and others in the great diversity of aromas that make wines famous. Analysis of grapevine reference genome has shown a remarkable expansion of certain families of genes linked to secondary metabolism in comparison with the other plants. In this work, I have analysed gene families coding for *cytochromes P450*, some of them being involved in the production of aromas, genes coding for stilbene synthases (*STS*), endo- β -1,3-glucanases and *NBS* type resistance genes involved in grapevine defences. My thesis intends to propose hypothesis to explain the structural organisation of these families and therefore better understand why some of these families are amplified in the grapevine genome.

Bioinformatic approaches have been used to study these different genes families. The *cytochromes P450* and R genes of *NBS* type were manually annotated to improve the knowledge of these families of genes. The expression of endo- β -1,3-glucanases, *STS* and *cytochromes P450* genes has been quantified using a large-scale transcriptomic approach. To this purpose, a tool has been developed during this thesis to estimate the level of genes expression from RNA-Seq data available in public databases. In the meantime, DNA resequencing data from 56 cultivars and grapevine species have been analysed to identify structural variations of CNV types within the genes with a *NBS* domain and the *STS* genes.

These works showed that the amplification of the gene families of interest was not specific to the reference genome but occurred at the scale of the *Vitis* genus, but also to highlighted structural variations in different genomes. Regarding the *STS* genes, blocks of duplication and more conserved and expressed genes were identified. For the genes with *NBS* domain, a clustered organisation has been highlighted with some clusters varying more than others in the studied genotypes. These works contribute to a better knowledge of gene families for efficient and durable defence against pathogens and optimal aromas synthesis in grapevine. This knowledge will benefit to breeding programs currently in progress at INRA Colmar.

Key words: grapevine, family of genes, comparative genomics, transcriptomics, CNV