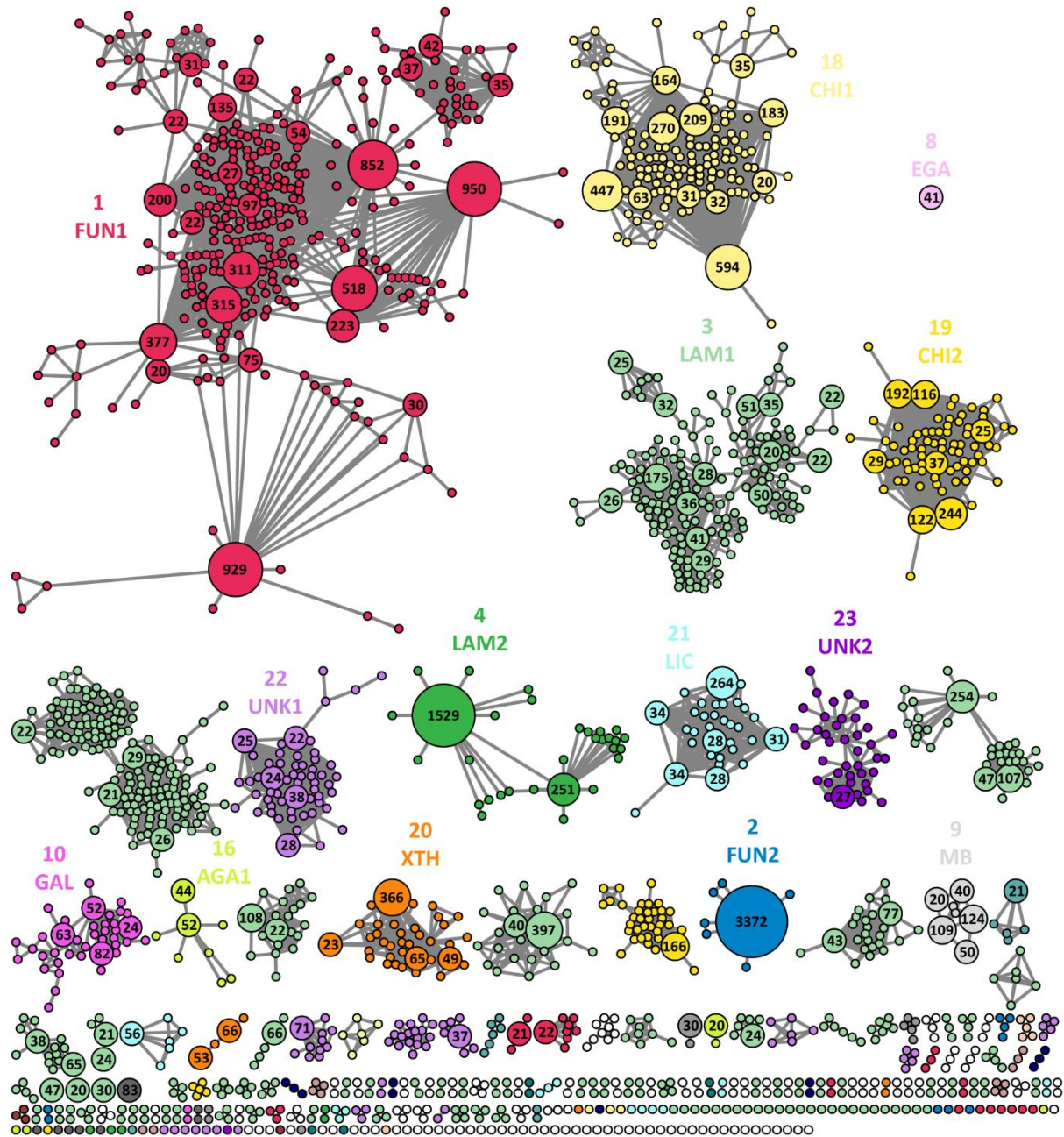


Subfamily delineation of the evolutionarily diverse Glycoside Hydrolase Family 16 (GH16) provides a roadmap for functional glycogenomics

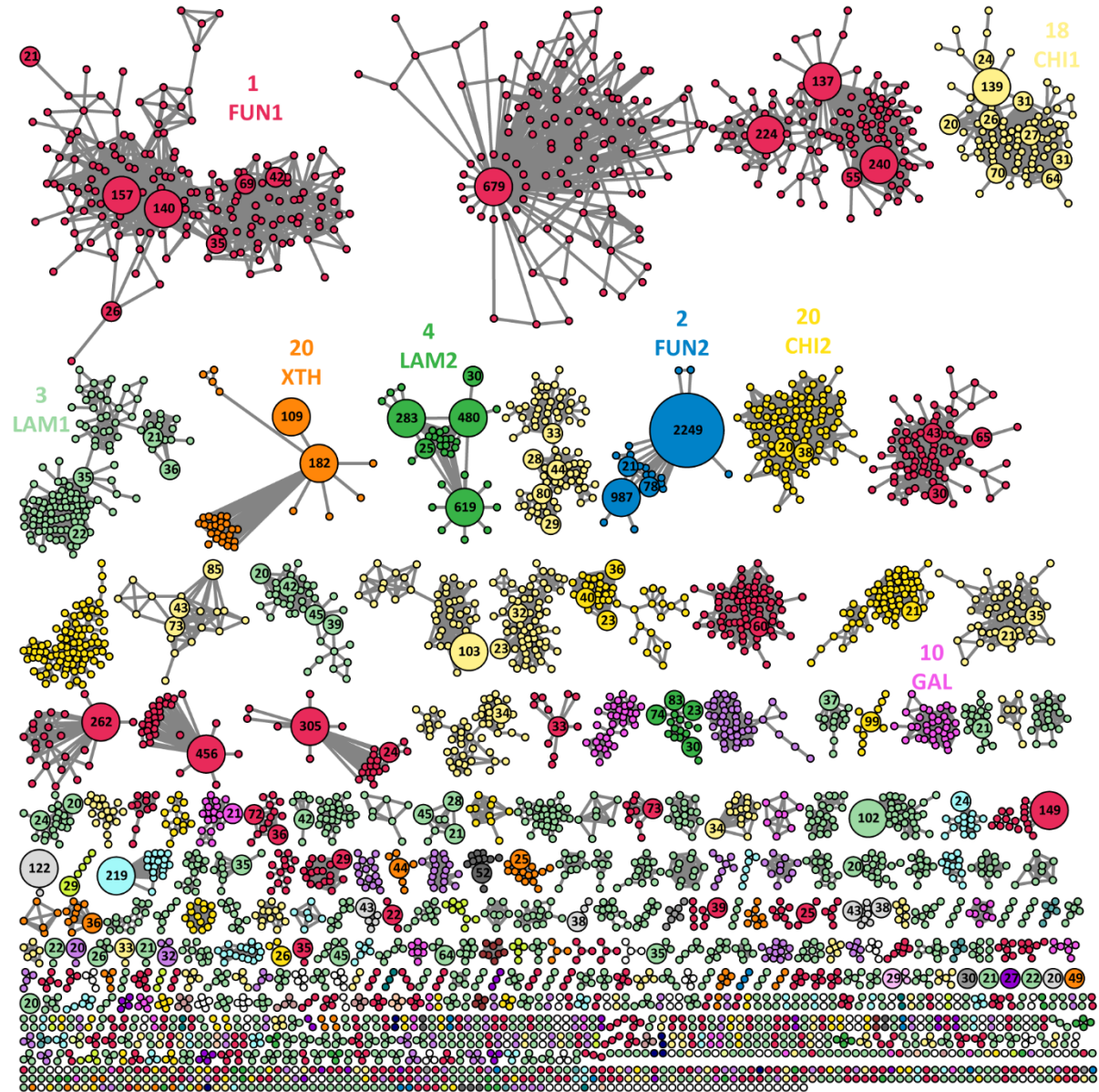
Alexander Holm Viborg<sup>1,=</sup>, Nicolas Terrapon<sup>2,3,=</sup>, Vincent Lombard<sup>2,3</sup>, Gurvan Michel<sup>4</sup>, Mirjam Czjzek<sup>4</sup>, Bernard Henrissat<sup>2,3,\*</sup>, and Harry Brumer<sup>1,5,6,7,\*</sup>

List of material included:

- Figure S1. SSN of 22,946 GH16 sequences at  $E = 10^{-85}$ .
- Figure S2. SSN of 22,946 GH16 sequences at  $E = 10^{-120}$ .
- Figure S3. Wall-eyed stereo representation of the superimposition of GH16 tertiary structures.

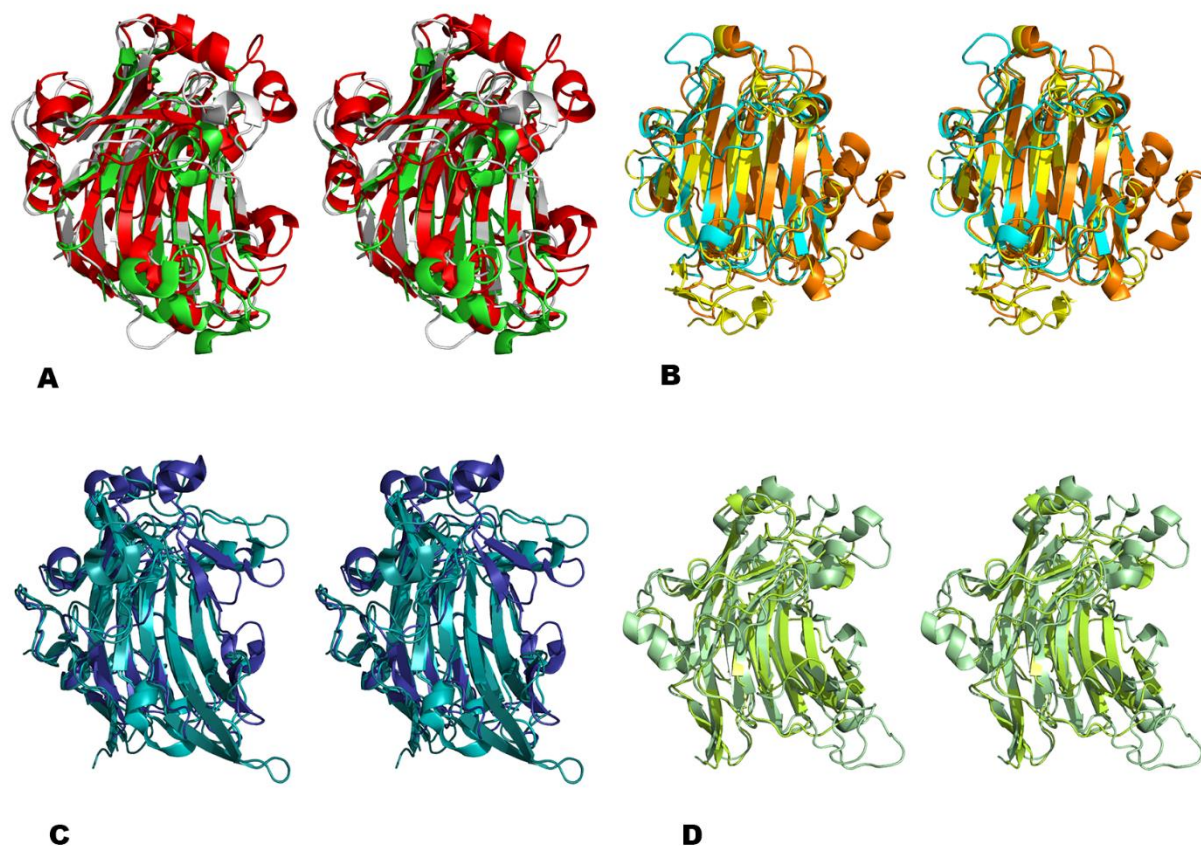


**Figure S1.** SSN of 22,946 GH16 sequences at  $E = 10^{-85}$ . Edges represent an  $E$ -value threshold below  $10^{-85}$  (*cf.* Figure 4). Meta-nodes represent merged sequences at increased threshold ( $E > 10^{-120}$ ); only meta-nodes containing 20 or more sequences are enlarged, with the number of merged sequences indicated. Each of the 23 defined subfamilies are indicated by a specific color (see Figure 2 for subfamily numbering and mnemonics). Not all subfamilies are labeled and only the biggest group of the segregated subfamilies is labeled. Non-classified sequences are indicated in white.



**Figure S2.** SSN of 22,946 GH16 sequences at  $E = 10^{-120}$ . Edges represent an  $E$ -value threshold below  $10^{-120}$  (cf. Figure 4). Meta-nodes represent merged sequences at increased threshold ( $E > 10^{-150}$ ); only meta-nodes containing 20 or more sequences are enlarged, with the number of merged sequences indicated. Each of the 23 defined subfamilies are indicated by a specific color (see Figure 2 for subfamily numbering and mnemonics). Not all subfamilies are labeled and only the biggest group of the segregated subfamilies is labeled. Non-classified sequences are indicated in white.





**Figure S3. Wall-eyed stereo representation of the superimposition of GH16 tertiary structures.** The 3D structures of members of neighboring subfamilies in the phylogenetic tree are superimposed using the secondary-structure-matching algorithm (SSM) implemented in COOT. The superimpositions highlight that despite the phylogenetic proximity of the members of these subfamilies, extensive differences are present, especially in the loops surrounding the active site groove. **A.** Superimposition of 2CL2 from GH16\_1 (red), 4CTE from GH16\_3 (light green) and 4PQ9 from GH16\_9 (light grey). **B.** Superimposition of 5NDL from GH16\_18 (light yellow), 2VH9 from GH16\_20 (orange) and 1GBG from GH16\_21 (cyan). **C.** Superimposition of 3JUJ from GH16\_11 (light steel), 4AWD from GH16\_12 (steel) and 5OCR from GH16\_17 (dark blue). **D.** Superimposition of 4ATF from GH16\_16 (lime green) and 6HY3 from GH16\_15 (pale lime green). The colors of the ribbon representations correspond to those of the respective group in the phylogenetic tree (Figure 5A).