**Figure 1**

Clustering of BP-Pol sequences. a) The first step of the clustering; SSN network with nodes representing individual proteins and edges representing pairwise alignment bit scores. Proteins are linked by edges if they have a pairwise score above 110. The resulting clusters are sorted according to number of protein members, with the largest cluster in the upper left corner. b) The second step of the clustering: HMM models were built for each SSN cluster and the HMMs were compared using HHblits. A network was built with nodes representing SSN clusters and edges representing HHblits scores. SSN clusters are linked by edges if they have an HHblits score higher than 160. The resulting clusters are referred to as superclusters and are sorted according to number of SSN clusters. There are two edges between nodes, when the HHblits score is above 160 in both directions. The size of the nodes represents the number of members in the SSN cluster. The 14 largest superclusters (>150 GenBank members) define CAZy families GT122 - GT135. Nodes are colored consistently according to their respective CAZy family in both panel a and b.

**Figure 2**

Level of conservation of stereochemical outcome of the reaction catalyzed in the various BP-Pol families. The bars represent the number of enzymes that are known to employ either retaining (making an axial bond) or inverting (making an equatorial bond) mechanisms in each of the new BP-Pol families.

**Figure 3**

Comparison of repeat unit sugars transferred by BP-Pols in GT125. The transferred repeat unit structures (in SNFG representation) are shown on a phylogenetic tree of BP-Pols in family GT125. There is an overall similarity between all the transferred sugars in the family and the similarity appears to correlate with the tree structure, ie. BP-Pol sequence similarity. In particular, the ends of the repeat units (+1 and -1 subsites) appear to be often conserved, whereas there is more variety in the central region where the enzyme does not interact with the sugar. Note that the +1 site corresponds to the non-reducing end of the depicted sugar structures and the -1 site corresponds to the reducing end. Notably, the family contains BP-Pols from distant taxonomic origin and that yet transfer similar repeat units.

**Figure 4**

An idealized representation of a BP-Pol. The donor is the growing glycan chain activated by Und-PP while the acceptor is a single repeat unit linked to Und-PP. The reaction is hypothesized to chiefly involve the sugar residues of the donor (subsites -2 and -1) and of the acceptor (subsites +1 and +2) that are proximal to the reaction center rather than residues and branches that are more distal (depicted in dashed lines). The reaction is represented by red arrows.

**Figure 5**

Glycan similarity of sugar repeat units polymerized by BP-Pols. All seed BP-Pols where the corresponding transferred oligosaccharide was known were included in the heatmap. A phylogenetic tree is shown for the polymerases in each CAZy family on the left. The glycan similarity scores are shown in a color scale of light blue (score value of 2 corresponding to identical matches at both –1 and +1 sites) to dark blue (score value of 5 corresponding to identical matches for at least three additional sequential positions). Horizontal lines separate the families. The darker colors close to the diagonal and within the families indicate specific substrate similarities in each family.

**Figure 6**

Relatedness of the new CAZy families and definition of clans. Inter-family HHblits bit scores are shown in a heatmap on a color scale from white (low similarity score) to dark blue (high similarity score). The HHblits scores depend on the direction of the alignment, and therefore the heatmap is not symmetrical. The inverting BP-Pols form two clans, GT-C1 which also contains the inverting SEDS (GT119) and the inverting O-Ligs (GT121) and GT-C3 containing only BP-Pols. The retaining BP-Pol families form one clan, GT-C2, which also contains the retaining ECA-Pol family (GT120).

**Figure 7**

Equivalent conserved residues in the clans. Conserved residues of each of the new CAZy families are shown on sequences of representative family members. Colored lines are shown between conserved residues from different families, which align in HHblits alignments and co-localize in structural superimpositions (Supplementary Figure 5-7). Transmembrane helices are shown in dark gray boxes, extracellular helices are shown in light gray boxes. The secondary structures were taken from the crystal structures for family GT119 and GT121 (6BAR and 7TPG respectively) and from AlphaFold models for all other families. The R210 in GT131 is either K or R in the family.

**Figure 8**

Structural superimpositions of members of different functional classes belonging to the same clans. a) Superimposition of O-Lig in cyan (GT121, PDB: 7TPG) and AlphaFold model of BP-Pol in pink (GT126, Genbank accession: AAM27615.1) (RMSD 5.3Å over 192 residues, sequence identity 20.8% over 485 residues) showing that the conserved Glu in the BP-Pol aligns with the conserved His in the O-Lig, which has been proposed to activate the acceptor21. b) Structural superimpositions of AlphaFold models of ECA-Pol in cyan (GT120, Genbank accession: ACH50550.1) and BP-Pol in pink (GT134, Genbank accession: CAI32772.1) illustrating structural similarity and co-localization of the conserved residues (RMSD 5.4Å over 360 residues, sequence identity 17.1% over 543 residue).

**Figure 9**

Oligosaccharide translation from IUPAC nomenclature to backbone (geometric) subunits for a trisaccharide consisting of one D-galactopyranose and two D-glucopyranose residues joined by intramolecular β1→3 and α1→4 bonds, respectively, and an intermolecular α1→3 bond formed in the polymerase reaction. (a) IUPAC nomenclature (b) Stereochemical projection highlighting backbone (thick grey line) and transfer bond (hatched line segments), and translated geometric subunits below. (c) Completed translation.