SUPPLEMENTARY MATERIAL 1.

The member databases themselves produce regular releases. PRINTS produces quarterly releases with 50 new fingerprints per release, resulting in 200 additional fingerprints per annum. At InterPro's conception Pfam had 2008 HMMs, and plan to reach a total of 5000 families by the end of 2002. In 2000 they produced 715 HMMs, in 2001 735 HMMs and aim to have produced 1700 additional HMMs by the end of 2002. For TIGRFAMs, the number of models has increased from 1109 in release 1.0 (2001) to 1415 in release 2.0 (beginning of 2002). The first release of PROSITE in 1989 contained just 60 entries, and today release 17.0 has 1501 signatures. Release 12.0 in 1994 saw the introduction of the first profiles into the releases, and since then they have produced an average of just over 100 new signatures per release (approximately per year).

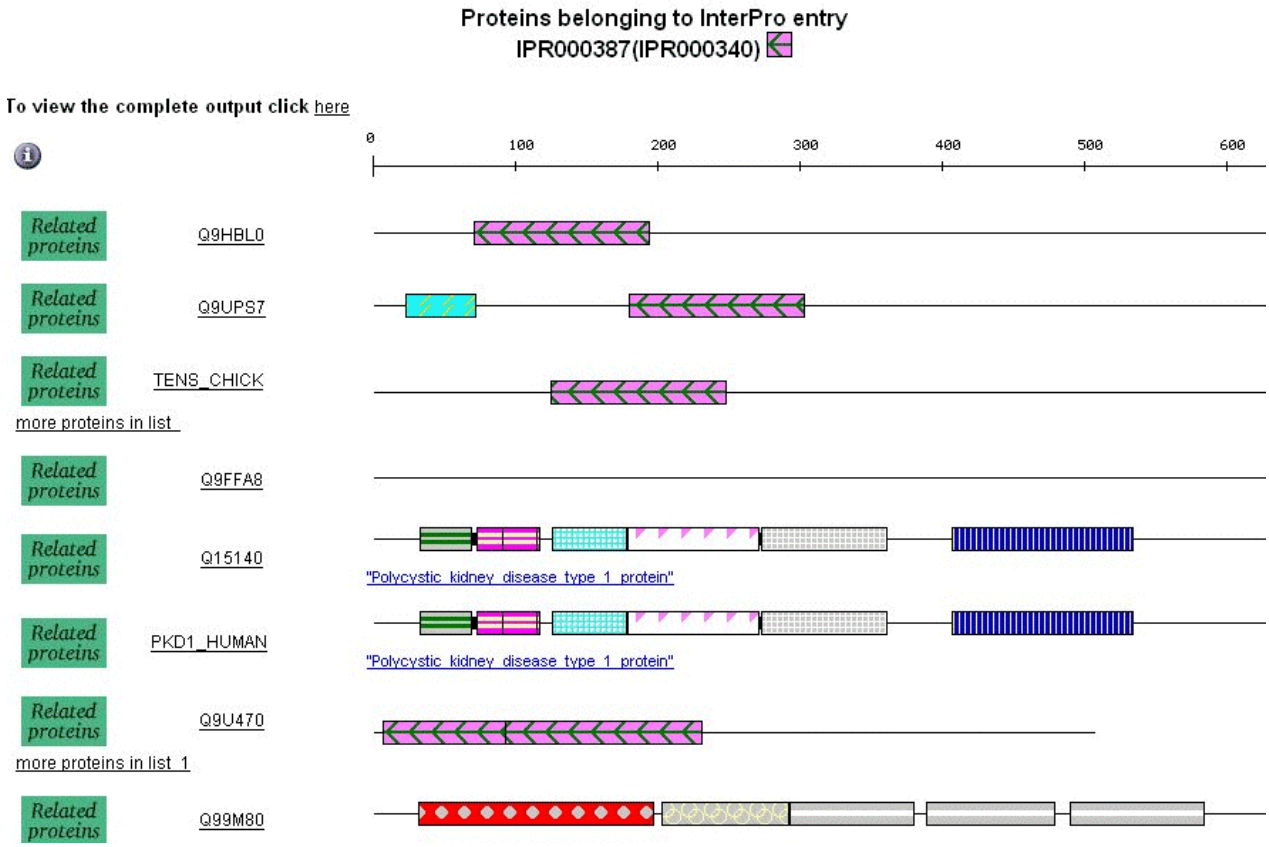
SUPPLEMENTARY MATERIAL 2.

Not all of the member databases strive purely for increased coverage of SPTR when considering new signatures. For example, PRINTS is primarily research driven, they focus on particular gene families, and the creation of hierarchical family discriminators in order to provide reliable estimates for the number of particular proteins in a given proteome. Traditionally they have concentrated on GPCRs, ion channels, peptidases, cytochrome P450s, tubulins, bacterial virulence factors, and a variety of other families. Nevertheless, PRINTS fingerprints have a coverage of 18.15% in a non-redundant SPTR, not including some of the fragment sequences present. Pfam on the other hand are coverage driven and strive to include the largest families they don't already have. The coverage of SPTR by PfamA families increased from 57.2% in 1998 to 72.5% in 2002. Increased coverage of SPTR by SMART,

currently 34.8% coverage by 641 domains, is a side effect of the research they carry out (9). TIGRFAMs focus mainly on prokaryotic genomes, but do have some models for eukaryotic protein families. PROSITE patterns and profiles cover 34.66% of protein sequences, and the motivation for new signatures is dependent on their importance for SWISS-PROT annotation. ProDom, with the nature of their methods of sequence clustering, have almost full coverage of SPTR (99.5% of non-fragmentary sequences in SPTR -36% of SPTR is comprised of fragment sequences), however not all ProDom entries are integrated into InterPro. The bigger ProDom families, particularly those with corresponding signatures from other member databases are integrated into InterPro with a higher priority.

SUPPLEMENTARY MATERIAL 3.

Figure 1. New graphical user interface for viewing protein matches of a particular InterPro entry. a.) Graphical view of representative list of proteins matching IPR000340, in which consensus domain boundaries have been computed for the domain line, and parent and children entries have been collapsed into one family line. This enables the family and domain composition information to be seen at a glance.



b.) From the "more proteins in list" link in view a.) it is possible to show all proteins sharing a common domain architecture. These protein sequences can then be retrieved or their alignments can be visualised.

- Graphics for list 9 : Get or Align proteins sharing the same architecture as **YNBD_ECOLI** :

YNBD_ECOLI 

YNBD_ECOLI	Q8X9S8	Q9I0U5		

- Graphics for list 10 : Get or Align proteins sharing the same architecture as **DUSC_HUMAN** :

DUSC_HUMAN 

DUSC_HUMAN	Q93VP1			

- Graphics for list 11 : Get or Align proteins sharing the same architecture as **O61722** :

O61722 

O61722	Q15197	Q22582	Q9NK63	