

# A Genomic Survey of HECT Ubiquitin Ligases in Eukaryotes Reveals Independent Expansions of the HECT System in Several Lineages

Xavier Grau-Bové<sup>1</sup>, Arnau Sebé-Pedrós<sup>1,\*</sup>, and Iñaki Ruiz-Trillo<sup>1,2,3</sup>

<sup>1</sup>Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Catalonia, Spain

<sup>2</sup>Departament de Genètica, Universitat de Barcelona, Catalonia, Spain

<sup>3</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

\*Corresponding author: E-mail: arnau.sebe@ibe.upf-csic.es.

Accepted: March 29, 2013

**Data deposition:** Protein alignments from this project have been deposited at <http://www.datadryad.org> under the accession doi:10.5061/dryad.mt620.

## Abstract

The posttranslational modification of proteins by the ubiquitination pathway is an important regulatory mechanism in eukaryotes. To date, however, studies on the evolutionary history of the proteins involved in this pathway have been restricted to E1 and E2 enzymes, whereas E3 studies have been focused mainly in metazoans and plants. To have a wider perspective, here we perform a genomic survey of the HECT family of E3 ubiquitin-protein ligases, an important part of this posttranslational pathway, in genomes from representatives of all major eukaryotic lineages. We classify eukaryotic HECTs and reconstruct, by phylogenetic analysis, the putative repertoire of these proteins in the last eukaryotic common ancestor (LECA). Furthermore, we analyze the diversity and complexity of protein domain architectures of HECTs along the different extant eukaryotic lineages. Our data show that LECA had six different HECTs and that protein expansion and N-terminal domain diversification shaped HECT evolution. Our data reveal that the genomes of animals and unicellular holozoans considerably increased the molecular and functional diversity of their HECT system compared with other eukaryotes. Other eukaryotes, such as the Apusozoa *Thecanomas trahens* or the Heterokonta *Phytophthora infestans*, independently expanded their HECT repertoire. In contrast, plant, excavate, rhodophyte, chlorophyte, and fungal genomes have a more limited enzymatic repertoire. Our genomic survey and phylogenetic analysis clarifies the origin and evolution of different HECT families among eukaryotes and provides a useful phylogenetic framework for future evolutionary studies of this regulatory pathway.

**Key words:** ubiquitination pathway, posttranslational regulation, multicellularity, last common ancestor of eukaryotes, Holozoa.

## Introduction

Proteins are the main structural and functional components of all cells. To efficiently respond to different environmental conditions, the protein levels need to be constantly regulated. The ubiquitination pathway is one of the most important post-translational mechanisms for regulating protein turnover and molecular cell dynamics (Rotin and Kumar 2009). It is based on the posttranslational modification of proteins by the ligation of ubiquitin, a 76 amino acid signaling peptide that is conserved across eukaryotes. This ubiquitin flag targets the proteins to a number of different outcomes, such as protein degradation, membrane sorting, and signaling functions (Rotin and Kumar 2009). The ubiquitination pathway involves the sequential

transfer of activated ubiquitin (Ub) from E1 (ubiquitin activating enzyme) to E2 (ubiquitin conjugating enzyme), and subsequently from E2 to E3 (ubiquitin ligase), which binds Ub to the protein of interest. E3 ubiquitin ligases transfer Ub to one or more Lys residues in the substrate by linking the C-terminal Gly of Ub with a Lys of the target protein (and/or a Lys of the Ub itself). Ubiquitination can occur in different forms (Mukhopadhyay and Riezman 2007): mono-ubiquitination (attachment of a single Ub to a single Lys), multi-ubiquitination (several Lys residues tagged with Ub) and polyubiquitination (addition of a Ub chain to a single Lys of the target protein). Typically, mono- and multi-ubiquitination are related to sub-cellular localization processes such as the secretory and endocytic pathways (Hicke 2001). Polyubiquitination, on the other

© The Author(s) 2013. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

hand, directs proteins to the 26S proteasome (a multiprotein complex consisting of 19S regulatory and 20S catalytic sub-complexes), which recognizes ubiquitinated proteins and degrades them; a common fate for misfolded or damaged proteins (Pickart and Fushman 2004).

To date, several studies have been carried out to resolve the evolutionary history of the ubiquitination pathway from a pan-eukaryotic point of view. These studies have, however, focused on the most conserved elements of the system, that is, the E1 (Burroughs et al. 2009) and E2 enzymes (Burroughs et al. 2008; Michelle et al. 2009; Ying et al. 2009), revealing that this pathway is ancient and widely distributed in all the considered eukaryotic lineages—as it is also the case for the ubiquitin proteins themselves (Burroughs et al. 2007).

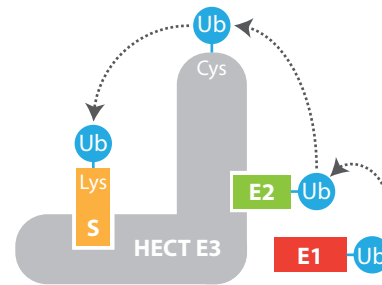
Conversely, most studies on E3 ubiquitin ligases have focused mainly on animals (Rotin and Kumar 2009; Marín 2010) and plants (Downes et al. 2003); and so little is known about the origin and evolution of these ligases within eukaryotes, and their relative importance in different eukaryotic lineages.

E3 ubiquitin ligases are of particular interest in evolutionary studies of the ubiquitination system, because they are way more diversified than E1 and E2 enzymes. The reason for this is that they are responsible for the specificity of the ubiquitination system, that is, they recognize, discriminate, and interact with the proper protein substrate (Rotin and Kumar 2009), and therefore are more functionally specialized. In fact, there are various groups of E3 enzymes according to their quaternary structure, their specific domain arrangements and the way in which they interact with E2 and the target protein. This includes, for instance, the HECT and RING ligases, and the CRL complexes. These proteins typically have a wide range of domain architectures involving specific protein–protein interaction motifs.

Indeed, the few eukaryotic genomes so far analyzed often encode many more E3 enzymes than E1 or E2. For example, there are more than 600 types of E3 in the human genome, whereas there are only two E1 proteins and approximately 30 E2 proteins (Schwartz and Ciechanover 2009).

HECT proteins are defined by the specific HECT domain, a C-terminal domain of approximately 350 amino acids that is essential for their Ub-ligase activity. The HECT domain is exclusive to HECT E3 ligases and is widespread among eukaryotes (Punta et al. 2012). HECT proteins directly intervene in the ligation process by forming an intermediate thioester bond between a highly conserved cysteine residue and Ub that binds Ub to the substrate (fig. 1) (Rotin and Kumar 2009).

Previous studies have devised a phylogenetic classification of animal HECTs (Marín 2010); however, there is little knowledge on the diversity of HECTs among all eukaryotes. Here, we perform a genomic survey of HECT ligases in eukaryotes and provide a useful evolutionary framework for future analyses. We also analyze the diversity of protein domain architectures of HECTs along the different eukaryotic lineages, as well as the putative relationship between the expansion of the



**Fig. 1.**—Schematic representation of Ub ligation to a protein substrate with a HECT ligase. The ligation process involves transferring the Ub from an activating enzyme (E1) to a transferase (E2) and then to the HECT ligase (E3). The E3 then ligates the Ub to a Lys residue of the substrate (S) with a thioester bond, involving a Cys residue in the HECT enzyme itself.

HECT-dependent ubiquitination system and the origin of multicellularity in several eukaryotic clades.

## Materials and Methods

### Taxon Sampling and Sequence Retrieval

HECT sequences were obtained from sequence data from complete genome sequences of 44 taxa, which represented all the recognized eukaryotic supergroups. Taxon sampling included 9 animals, 5 unicellular Holozoa, 8 Fungi, 1 Apusozoa, 3 Amoebozoa, 3 plants, 5 unicellular algae, 3 Heterokonta (1 being multicellular), and 11 other unicellular Bikonta (see [supplementary table S1, Supplementary Material](#) online). HECT amino acid sequences were retrieved with a HMMER search, using the HMM profile of the Pfam HECT domain entry (PF00632) as a query, the default parameters and an inclusive *E* value of 0.05. The search yielded 744 sequences (see [supplementary fig. S3, Supplementary Material](#) online).

### Protein Alignment, Manual Edition, and Data Curation

The retrieved sequences were aligned using Mafft (Katoh et al. 2002) L-INS-i algorithm (optimized for local sequence homology [Katoh et al. 2005]). The alignment was further edited manually and hits fulfilling one of the following conditions were removed: 1) incomplete sequences with more than 99% of sequence similarity with a complete sequence from the same taxa, and 2) sequences that showed extreme long branches in the preliminary maximum likelihood (ML) trees. The final alignment was carried out based on the HECT domain alone using the Mafft G-INS-i algorithm (for global homology).

### Phylogenetic Analyses

The phylogenetic trees of eukaryotic HECTs were inferred from both ML and Bayesian inference (BI) analyses, using

the LG evolutionary model with a discrete gamma distribution of among-site variation rates (four categories) and a proportion of invariable sites, which constituted the best model for this data set, according to Prottest (Abascal et al. 2005).

ML trees were estimated with RAxML 7.2.6 (Pthreads version [Stamatakis 2006]) and the best tree from 100 replicates was selected. Bootstrap support (BS) was calculated from 500 replicates. BI trees were estimated with Phylobayes 3.3 (Lartillot et al. 2009), using two parallel runs for 500,000 generations and sampling every 100. Bayesian posterior probabilities (BPPs) were used for assessing the statistical support of each bipartition.

### Domain Architecture Analysis

The N-terminal domain architecture of all retrieved sequences was inferred by performing a Pfam scan (Punta et al. 2012), using the gathering threshold as cut-off value. The domain information of each protein was used to 1) assess the reliability of each sequence of the initial data set, 2) help define protein families according to its architectural coherence, and 3) assess the level of functional and architectural diversification of HECT proteins across the eukaryote lineages. Additional information about some previously uncharacterized domain architectures was obtained from the bibliography and verified using manual protein alignments. The pattern of acquisition of new domains at the N-terminus of HECT proteins across the eukaryote tree of life was inferred using a strict parsimony approach based on phylogenetic information from BI and ML trees.

### Classification Criteria

The classification of the HECT proteins is based on two hierarchical categories: 1) protein families, which contain all proteins from orthologous genes with high nodal support, and 2) protein classes with one or more families, which are wider groups of phylogenetically related families that descend from one of the HECT proteins that have been inferred to exist in the last eukaryote common ancestor (LECA). Protein families sometimes share a common domain architecture, and therefore the domain content of each protein was used as an additional, conditional criterion to define some families. The pattern of gain and loss of families was inferred by strict parsimony based on phylogenetic information from BI and ML trees.

## Results and Discussion

### The Evolutionary Origin of HECT E3 Protein Family

Our phylogenetic analyses recovered six pan-eukaryotic clades of HECT proteins, defined as classes I to VI (figs. 2 and 3). Assuming the leading hypothesis that the root of eukaryotes lies between Unikonta and Bikonta (Stechmann and Cavalier-Smith 2002; Derelle and Lang 2012), our data imply that the last eukaryotic common ancestor had at least six HECTs that

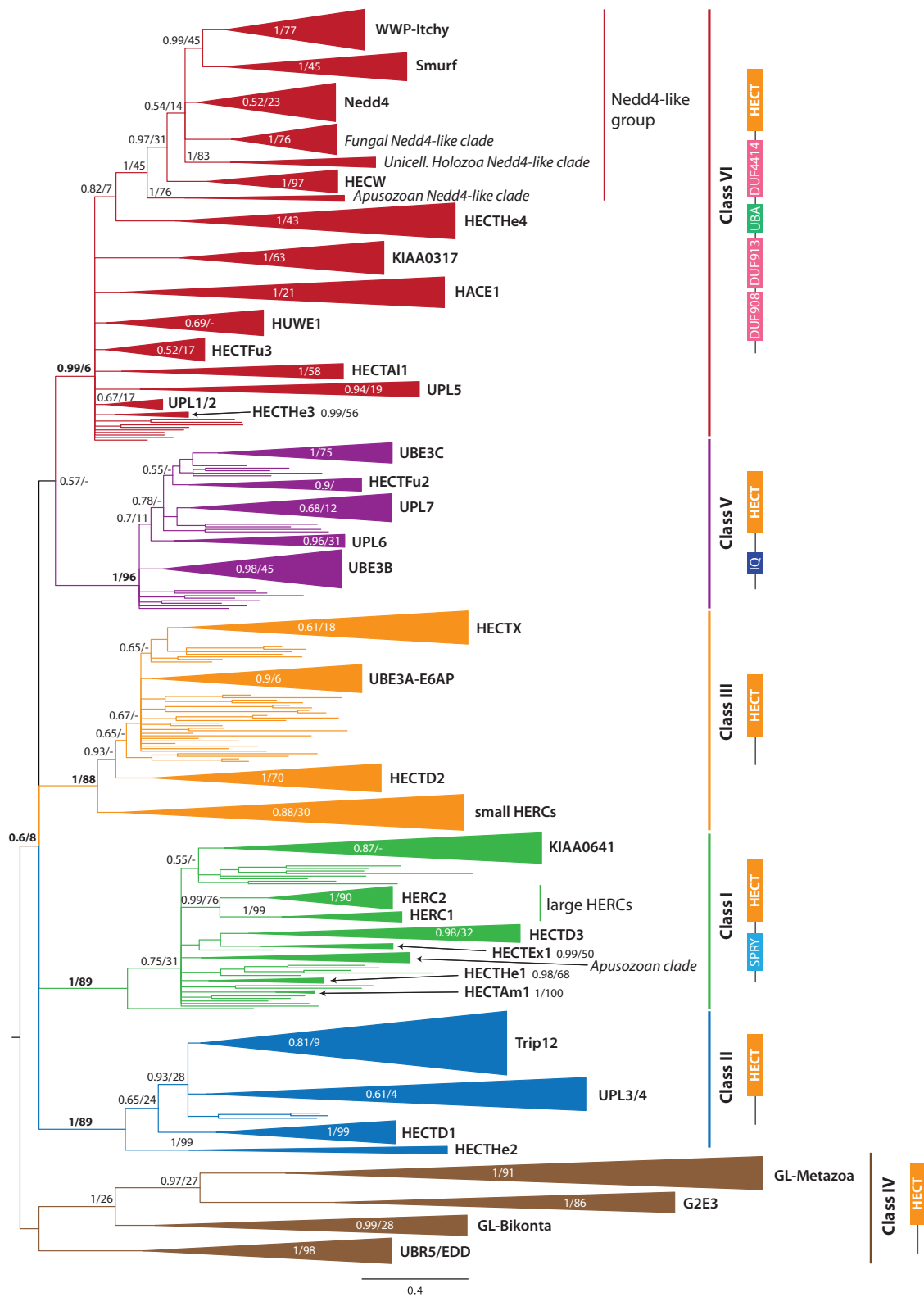
remain present in diverse eukaryotic lineages. In turn, these six main classes are divided into 35 distinct HECT families that are specific to certain eukaryotic lineages (fig. 3). This scenario remains the same if the alternative “Excavate-first” hypothesis of the root of the eukaryotes is considered (Rodríguez-Ezpeleta et al. 2007).

The diversification of each class involves many gene duplication events and secondary losses (fig. 4), as well as the acquisition of new accessory domains. Our data show that the protein domain architecture is quite diverse as a result of domain rearrangements and the acquisition of new domains at the N-terminal region (fig. 3).

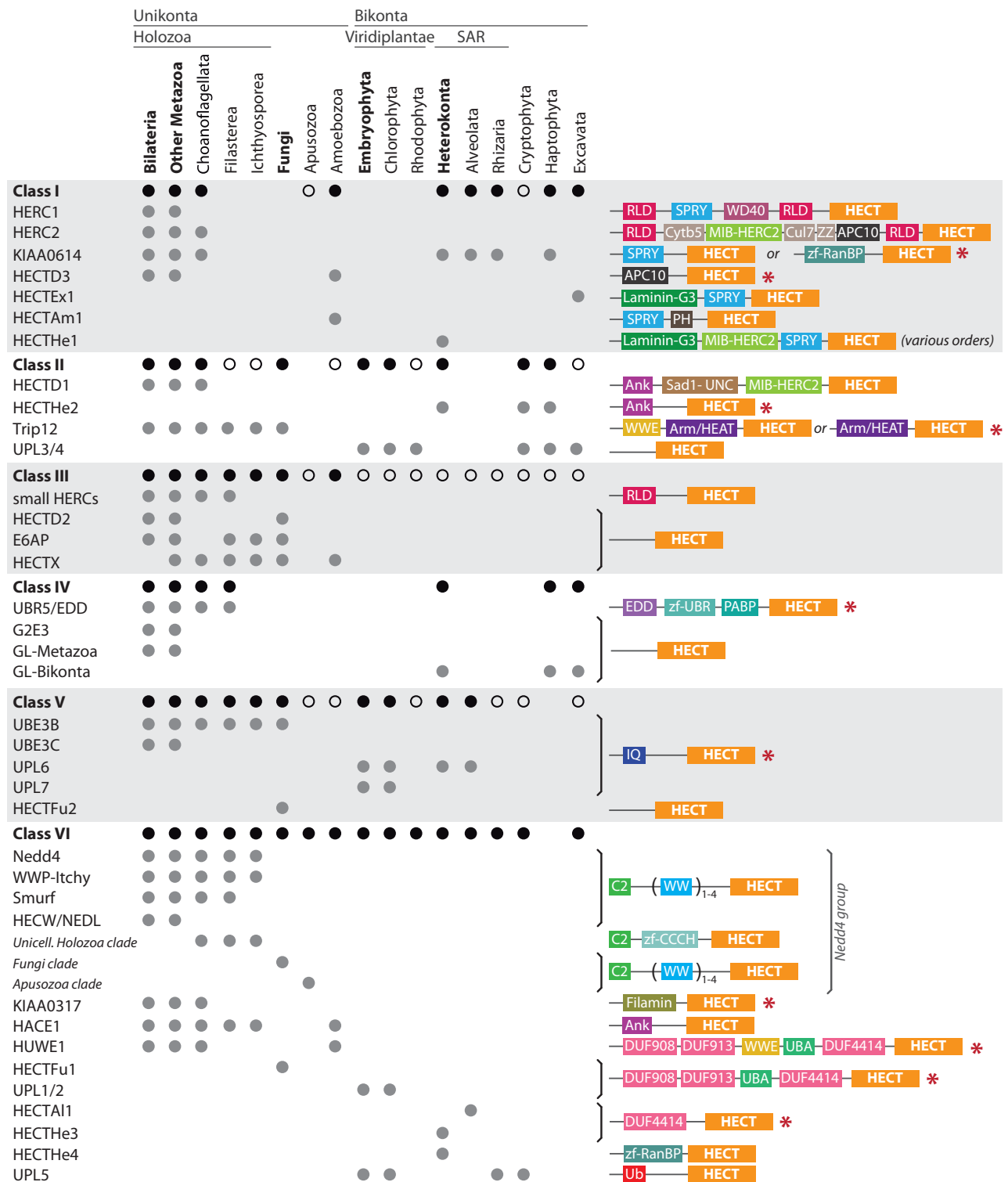
Remarkably, domain fusions at the C terminus have not been detected in any of the analyzed organisms. This might be explained by the fact that the catalytic activity of the HECT domain strongly depends on its tertiary structure: all HECTs are organized in two structurally distinct lobes (N-lobe and C-lobe, where HECT is located) that can adopt a limited range of three-dimensional conformations (Huang et al. 1999; Verdecia et al. 2003; Rotin and Kumar 2009). This tertiary structure is functionally relevant (and therefore constrained) because it defines the position of the catalytic cysteine residue with respect to the E2 enzyme and the ubiquitination substrate during the ligation process (Verdecia et al. 2003). It also determines the way in which the ubiquitin chain elongation occurs (Maspero et al. 2011).

Assuming the “Unikont–Bikont split” hypothesis on the root of eukaryotes (Stechmann and Cavalier-Smith 2002; Derelle and Lang 2012), the analysis of protein domain architectures reveals class-specific N-terminal domain arrangements that are pan-eukaryotically distributed in classes I (SPRY), V (IQ), and VI (DUF908, DUF913, UBA, and DUF4414), whereas the founding proteins of classes II, III, and IV (Rodríguez-Ezpeleta et al. 2007), a similar scenario emerges, except for the ancestral IQ (class V), DUF908, DUF913, and UBA domains (class VI), which are not recovered. However, DUF4414 (class VI) still appears to be present in the LECA.

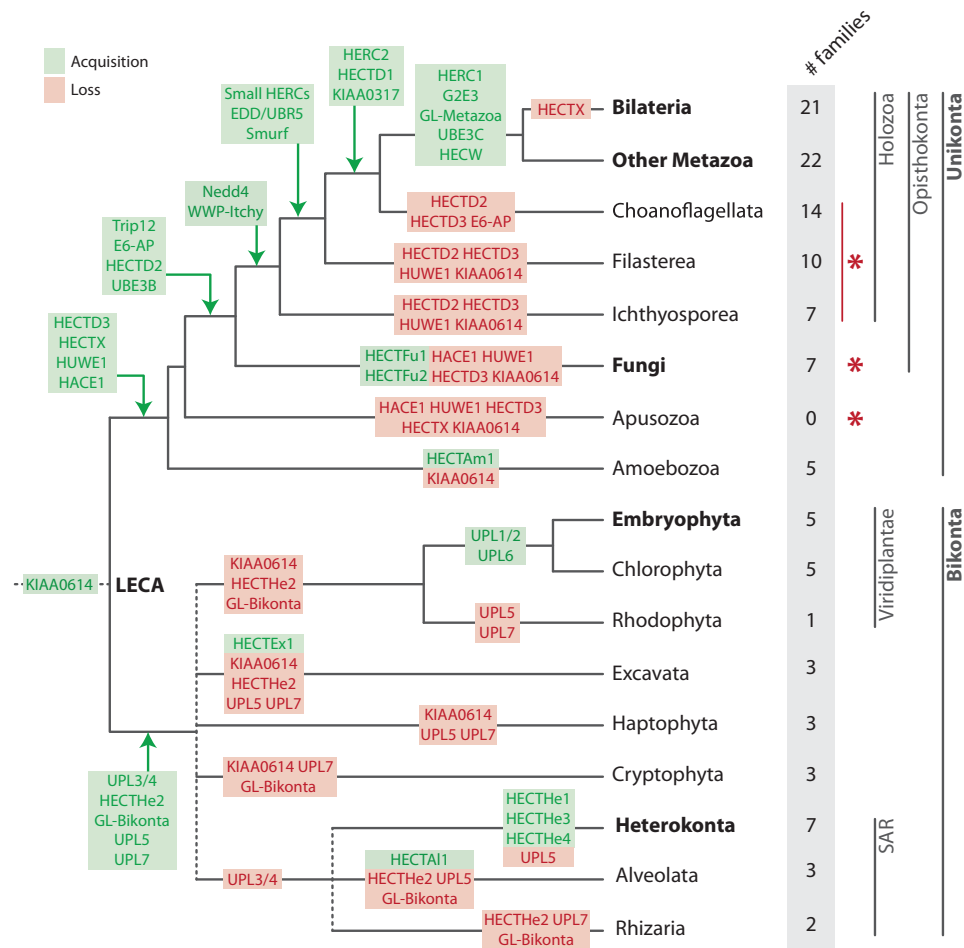
The syntax of N-terminal domain architectures in HECTs is mainly based on protein recognition motifs (IQ, WW, Ankyrin repeats, zinc fingers, etc.) that enable HECTs to specifically ubiquitinate certain substrates. Domains involved in targeting the HECT enzyme to certain molecules are also common, such as C2 (lipid binding), Laminin-G3 (complex sugar binding), and PABP (mRNA polyadenylate binding). Some of these motifs are especially “promiscuous” and have been independently gained several times thorough HECT evolution (for instance, ubiquitin-binding UBA and protein-binding domains such as WWE, SPRY, RCC1-like domain [RLD], Ankyrin, and MIB-HERC2) (fig. 5; details discussed later). Despite the generally conserved syntax of HECT N-terminal architectures, rare domains with no clear function exist on some uncharacterized HECTs. It is expected that the discovery of such unusual HECTs



**Fig. 2.**—BI phylogenetic tree of HECT proteins inferred from an alignment of the HECT domain (220 amino acid positions). Colored clades indicate classes; collapsed clades indicate families (in regular text) and other clades of interest (italics). Nodal labels indicate BPP and 500-replicate ML BS values, respectively. Dashes indicate that the node is not recovered. Six pan-eukaryotic classes can be distinguished, with 35 families within these. For each class, the putative ancestral N-terminal architecture is shown. Complete BI and ML trees are shown in [supplementary figures S1 and S2, Supplementary Material](#) online.



**Fig. 3.**—Schematic representation of HECT E3 classes and families with their archetypical domain architectures and presence/absence information in the sampled taxonomic groups. Filled circles signify that a given taxon has representatives in a certain family; empty circles signify that proteins of a given taxon are part of a class but cannot be classified into any described family. For each family, its typical architecture is shown on the right. Red asterisks indicate that a single HECT domain is also found in some sequences of that family. For more information on the architecture of each analyzed protein (supplementary figs. S1 and S2, Supplementary Material online). Taxa containing multicellular organisms are shown in bold (namely, Bilateria, other Metazoa, Fungi, Embryophyta, and Heterokonta).



**Fig. 4.**—Pattern of acquisition and loss of HECT families across the eukaryote tree of life, assuming that LECA is placed at the Unikonta–Bikonta split. Green and red boxes indicate acquisition and loss of families, respectively. The number of families with representatives from each clade is shown at the right. Red asterisks indicate taxonomic groups that have one additional HECT clade that cannot be classified as part of any family (but are still known to be part of the Nedd4 group). Taxa containing multicellular organisms are shown in bold.

will increase when more and more genomes are taken into account in future similar surveys.

### Classification of Eukaryotic HECT E3 Ligases

We have classified the different eukaryotic HECTs in different classes and families, according to the topology obtained by the phylogenetic analyses. A description of the main characteristics of each class and family is given in the following section.

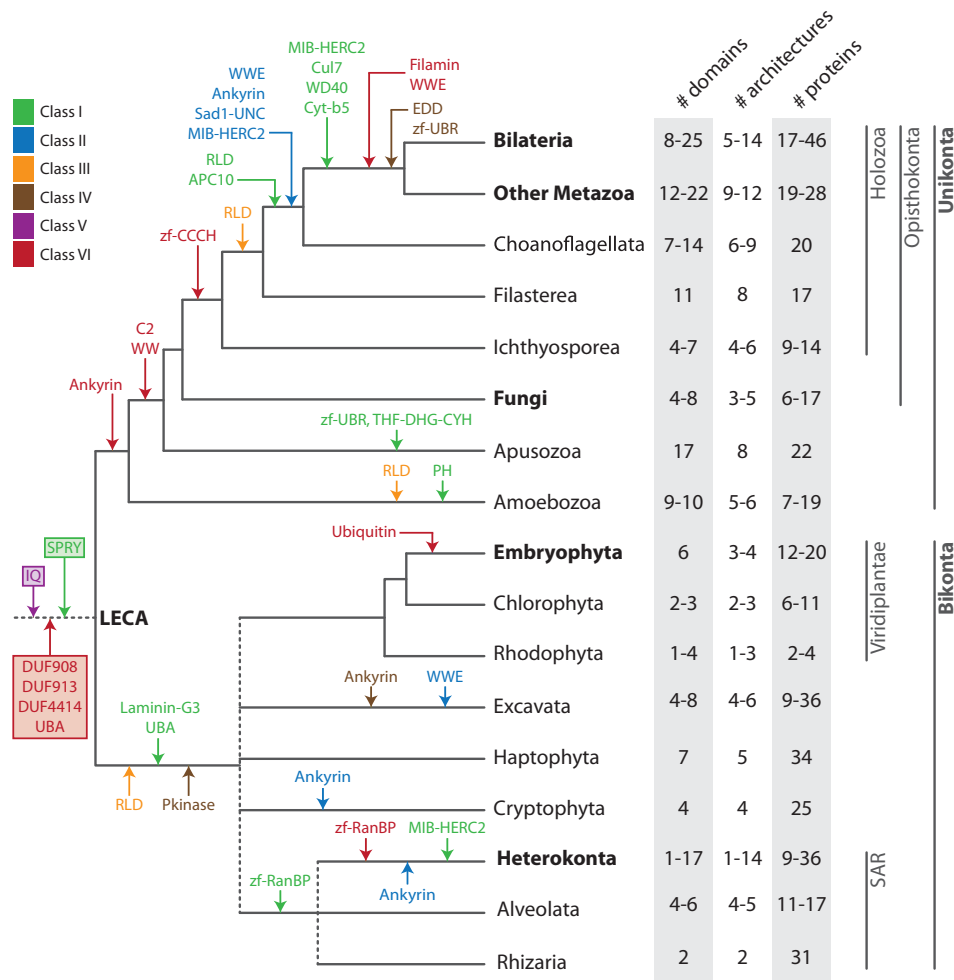
#### Class I: Large HERCs and Related Families

Class I contains seven protein families: HERC1, HERC2 (both known as large HERCs), KIAA0614, HECTD3, HECTXex1, HECTAm1, and HECTHe1 (figs. 2 and 3). The monophyly of class I is supported by a BPP of 1.0 and a BS value of 89% (fig. 2). Large HERCs were previously thought to be related to the family of small HERCs (class III in our tree), because they

shared the RLD (Hadjebi et al. 2008), but our data corroborate that these families are paraphyletic and the domains have been independently acquired (Gong et al. 2003; Marín 2010).

HERC1 is an animal-specific family that has been lost in Arthropoda (*Daphnia pulex* and *Drosophila melanogaster*) and Hemichordata (*Saccoglossus kowalevskii*). HERC1 proteins have a specific domain architecture consisting of HECT, two RLDs, SPRY, and a variable number of WD40 repeats. In some cases, there is also a UBA domain. In humans, HERC1 binds to clathrin heavy chain and has GEF activity on ARF1, a GTPase involved in membrane trafficking in the Golgi apparatus (Rosa and Casaroli-Marano 1996). HERC1 also ubiquitinates the tumor suppressor TSC2 (involved in the tuberous sclerosis complex disease and perhaps in membrane trafficking [Chong-Kopera et al. 2006]).

The HERC2 family, which appears as a sister group to HERC1, is closely related to HERC1 and includes proteins from both Metazoa and Choanoflagellata. In mammals,



**Fig. 5.**—Pattern of appearance of new domains in HECT proteins within the tree of eukaryotes. Domains are color coded by class (see legend). Each label indicates the acquisition of a new domain, considering that the appearances of the same domain in different classes are independent acquisitions. Note that the domain list is nonextensive, because it only takes into account domains from proteins that have been phylogenetically classified. Domains in boxes at the LECA indicate the inferred domain content of the ancestral protein that gave rise to classes I (SPRY), V (IQ domain), and VI (DUF908, DUF913, DUF4414, and UBA). The ancestral architectures of the classes III, II, and IV LECA proteins were a single HECT domain. Note that RLD, WW, Ankyrin, MIB-HERC2, and UBA have been convergently acquired several times in the evolution of different classes. The adjacent table contains information about the number of HECTs in each clade, the number of distinct domains in these proteins and the number of distinct domain architectures (shown as intervals; see [supplementary table S2, Supplementary Material](#) online, for more detail). Taxa containing multicellular organisms are shown in bold.

HERC2s ubiquitinate and target BRCA1 (breast cancer suppressor) for degradation (Wu et al. 2010). They have a complex domain architecture with two RLDs and several protein recognition motifs: Cyt-b5 (Ozols 1989), MIB-HERC2 (also present in RING E3 Mib2 [Itoh et al. 2003]), Cul7 (present in RING E3s Cul7 [Kaustov et al. 2007]), ZZ, and APC10. This architectural diversification occurred at the origin of the Metazoa, since the choanoflagellate homologs from both *Monosiga brevicollis* and *Salpingoeca rosetta* have simpler architectures (RLD repeats and RLD, APC10, and SPRY domains, respectively).

The KIAA0614 family is a pan-eukaryotic family with homologs in Metazoa, Choanoflagellata, Heterokonta, Alveolata,

Rhizaria, and Haptophyta. Some proteins have a SPRY domain, while proteins from *Phytophthora infestans* and *Tetrahymena thermophila* have an extra zf-RanBP.

The HECTD3 family contains animal proteins (bearing an APC10 domain) and a homolog from *Acanthamoeba castellanii*. Human HECTD3 ubiquitinates some proteins involved in neural development and brain function, such as Syntaxin-8 (Zhang et al. 2009) and Tara—which is also a regulator of cell growth, cytoskeletal actin reorganization and cell motility (Yu et al. 2008).

HERC2 and HECTD3 are the only HECT families with APC10 domains, and they both are exclusive to animals and choanoflagellates. APC10 domain is also found in the RING E3 APC/C

complex, which takes part in cell cycle control by regulating mitosis (Jin et al. 2008). In this context, APC10 is responsible for the regulation of substrate binding (Peters 2002).

The other families within this class (i.e., HECTEx1, HECTAm1, and HECTHe1) are named after their taxonomic content (Excavata, Amoebozoa, and Heterokonta) and are defined by their distinctive domain arrangements. For instance, HECTAm1 contains PH and SPRY motifs, and HECTHe1 and HECTEx1 have Laminin-G3 (capable to reversibly bind to specific complex sugars, an exclusive feature of these two families) and SPRY domains. Also, class I contains a clade with *Thecanomas trahens* proteins bearing various protein recognition domains that seem to have been independently acquired (fig. 2).

The SPRY domain is exclusive to class I HECTs and is present in most of its families, which suggests that it could have existed in the ancestral LECA protein that gave rise to this class. It has been reported that SPRY plays a role in the recognition of ubiquitination substrates (Nishiya et al. 2011).

### Class II

The well-supported class II (BPP = 1.0; BS = 89%) is composed of four protein families: HECTD1, HECTHe2, UPL3/4, and Trip12 (figs. 2 and 3).

The HECTD1 family contains sequences from Metazoa and Choanoflagellata. They have a distinctive protein domain arrangement containing Sad1-UNC, MIB-HERC2 domains and, in some cases, Ankyrin repeats. Human HECTD1 polyubiquitinates Hsp90, a chaperone that controls cell motility, which is essential in brain development (Sarkar and Zohn 2011). The HECTHe2 family also contains proteins with Ankyrin repeats and is specific to Heterokonta, Cryptophyta, and Haptophyta. Their functions are still unknown.

Trip12 (also known as ULF) includes proteins from animals, unicellular Holozoa and Fungi. Animal Trip12s are defined by two protein recognition domains: HEAT repeats, which are Armadillo-like motifs that recognize ubiquitin degradation signals in E3s substrates (Tewari et al. 2010); and WWE, which recognizes the Ankyrin motif of Notch and ligand-binding domains of other proteins (Aravind 2001). Fungal Trip12s also have HEAT/Armadillo repeats with a similar function, for example, the yeast Ufd4 HECT (Tewari et al. 2010).

Trip12 activity hampers tumor suppression in humans by preventing the p53 response to oncogenic events: it promotes the degradation of ARF, an inhibitor of the RING E3 Mdm2 (which in turn targets p53 for degradation [Brooks and Gu 2006]). Trip12 also targets p16 (a murine negative cell cycle regulator during embryogenesis) to degradation (Kajiro et al. 2011).

The UPL3/4 family includes homologs from several Bikonta clades (Viridiplantae, Excavata, Cryptophyta, Haptophyta, and Rhodophyta). Some Viridiplantae proteins also have Armadillo repeats, which have been predicted to recognize nuclear

localization signals (Downes et al. 2003). *Arabidopsis* UPL3 polyubiquitinates some unknown regulator of trichome development (Downes et al. 2003); and both UPL3 and UPL4 collaborate in the regulation of Gibberellin cell signaling (Coates 2008). However, concrete substrates remain elusive.

### Class III: Small HERCs, E6AP, and Other Families

Class III (BPP = 1.0; BS = 88%) includes small HERCs, HECTD2, E6AP (all of them named after the human proteins within them), and HECTX (Marín 2010) composed of Unikonta proteins. However, class III also includes proteins from Bikonta species (Viridiplantae, SAR, Cryptophyta, Haptophyta, and Excavata) that cannot confidently be assigned to any family, branching in an unclear position related to HECTD2, E6AP, and HECTX, but with low nodal supports.

The family of small HERCs includes proteins from animals, Choanoflagellata and Filasterea clades. It embodies human proteins HERC3, 4, 5, and 6, that is, the remaining HERC proteins that were formerly considered to be closely related to large HERCs 1 and 2 (see class I). So, any a priori functional or evolutionary similarities between these families need to be re-assessed. For instance, in contrast to large HERCs, the RLD motifs from small HERCs do not act as guanine nucleotide exchange factors (Rotin and Kumar 2009).

Indeed, convergent acquisition of RLD domains seems to be a common event in HECT evolutionary history: they are also present in several non-holozoan "HERC-like" proteins that cannot be assigned to any specific family (*A. castellanii*, *Toxoplasma gondii*, *Ectocarpus siliculosus*, *Cyanidioschyzon merolae*, and *Emiliana huxleyi* from class III; and *P. infestans* from class I). RLD domains intervene in a wide variety of cellular processes (RNA processing and transport, RNA mating, imitation of mitosis, chromatin condensation, guanine-nucleotide-exchange factor, protein recognition in DNA binding, and ubiquitination), which could explain their high "promiscuity."

Human small HERCs have important functions. For example, HERC3 binds Ub, PLIC1, or PLIC2 (Ub-like proteins) to endocytic proteins, thus regulating vesicular transport (Cruz et al. 2001). HERC4 is essential for spermatogenesis in mice (Cruz et al. 2001), and HERC5 is involved in the immune response related to interferon signaling pathways and polyubiquitinates I $\kappa$ B (inhibitor of the pro-inflammatory transcription factor NF- $\kappa$ B) (Kroismayr et al. 2004; Dastur et al. 2006).

The E6AP family (also known as E3A or UBE3A) includes the human protein E6AP (one of the first described HECTs), as well as proteins from animals, *Capsaspora owczarzaki*, *Sphaeroforma arctica*, and *Mortierella verticillata*, although the latter has poor nodal support. Human E6AP is known for its role in the inactivation of tumor suppressor p53 through proteasomal degradation (Scheffner 1998). E6AP is a good example of complex interplay between E3, in which different E3s have different antagonistic roles. For instance, human E6AP is polyubiquitinated by UBR5/EDD (another HECT E3,



discussed later) (Tomaic et al. 2011), as well as being enhanced (in an ubiquitin-independent manner) by HERC2 (Kühnle et al. 2011).

The HECTD2 family is an Opisthokonta-specific family that includes sequences from animals and Fungi, but not from unicellular Holozoa. HECTD2 proteins have a single HECT domain. Murine and human HECTD2 are known to intervene in protein degradation in neurodegeneration processes (Lloyd et al. 2009).

HECTX contains proteins from Cnidaria and Placozoa proteins, as well as from Filasterea, Fungi, and Amoebozoa. Thus, the lack of HECTX in bilaterians genomes is probably due to a secondary loss.

#### Class IV

Class IV includes four families: UBR5/EDD, G2E3, GL-Metazoa, and GL-Bikonta. The latter three are extremely divergent at the sequence level (figs. 2 and 3). The nodal support for this class is weak (fig. 2), but both Bayesian and ML analyses recovered the clade. In contrast, the nodal support for all of the families, except GL-Bikonta, is very good (BPP = 1.0 and BS = 99–100%).

The UBR5/EDD family includes proteins from animals (which have an EDD domain for binding ubiquitin, a zf-UBR protein recognition motif and a PABP domain) and architecturally simpler homologs from the choanoflagellate *Sal. rosetta* and the filasterean *Cap. owczarzaki*. Human EDD and *Dro. melanogaster* HYD act as general tumor suppressors by ubiquitinating E6AP (Tomaic et al. 2011), which increases p53 levels and induces cell senescence (Smits 2012). EDD and HYD also ubiquitinate TopBP1 (a topo-isomerase that intervenes in DNA damage response [Honda et al. 2002]) and negatively regulate Hh (hedgehog pathway) and Dpp (decapentaplegic pathway) expression, two crucial elements in the *Drosophila* eye disc development process (Lee 2002).

The G2E3, GL-Metazoa, and GL-Bikonta families are composed of proteins with a highly divergent HECT domain, with different domain arrangements that could confer them their own functional specificities. For instance, some proteins from *Naegleria gruberi* and *E. siliculosus* (GL-Bikonta) have unusual protein kinase domains of unknown function; and human and murine G2E3s have a non-functional HECT domain and three unconventional RING/PHD-like zinc fingers, two of which have been proved to have ubiquitin ligase activity (Brooks et al. 2008). None of these zinc fingers has been clearly classified as either PHD or RING motifs, although Pfam identifies the noncatalytically active one as a PHD-like zf-HC5HC2H domain (which is consistent with the fact that PHD domains are unable to act as ubiquitin ligases [Scheel and Hofmann 2003]). The lack of functional constraints on the HECT sequence would explain its divergence from other HECT proteins.

The most parsimonious explanation for the evolution of class IV is that an ancestral LECA gene underwent a

duplication that gave rise to 1) the holozoan EDD family (secondarily lost in Bikonta species), and 2) a fast-evolving group, including the G2E3, GL-Metazoa, and GL-Bikonta families.

#### Class V

Class V (BPP = 1.0; BS = 96%) contains five families with proteins from Unikonta and Bikonta: UBE3B, UBE3C, HECTFu2, UPL6, and UPL7 (figs. 2 and 3). Except for HECTFu2, proteins belonging to this class have an exclusive IQ domain that could have been present in the ancestral protein that gave rise to class V. IQ typically binds to calmodulin and is also present in proteins that interact with GTP regulatory and cell cycle proteins, receptors, and channel proteins (Rhoads and Friedberg 1997).

UBE3B is an Opisthokonta-wide family in which an IQ domain is present in some proteins from animals, Filasterea (*Cap. owczarzaki*) and Fungi (*M. verticillata*). Proteins from the animal family UBE3C also have an IQ domain. UBE3B is thought to play a role in the oxidative stress response in humans and *Caenorhabditis elegans* (Oeda et al. 2001), and UBE3C plays an undetermined role in inflammatory responses in the human airways, probably related to I $\kappa$ B ubiquitination (Pasaje et al. 2011).

The HECTFu2 family, defined here for the first time, is specific to Fungi and their proteins do not bear any particular N-terminal protein domain architecture. It has no known substrates.

The UPL6 and UPL7 families conform to two independent clades, both consisting of Embryophyta and Chlorophyta proteins. UPL7 also contains proteins from Alveolata and Heterokonta. Again, IQ domains are found in Embryophyta and Chlorophyta sequences from UPL7 and Embryophyta sequences from UPL6. Contrary to previous studies (Gong et al. 2003), we did not recover a sister-group relationship between UPL6 and UPL7.

#### Class VI: Nedd4-Like, HUWE1, HACE1, and Other Families

Class VI is a wide group that includes 13 families plus three unclassified clades (figs. 2 and 3). The Bayesian analysis provides a good nodal support for this class (BPP = 0.99), but the clade is not statistically supported by ML.

The Nedd4-like group contains all families with C2 and WW domains: HECW/NEDL (with 1–2 WWs; specific to animals) Nedd4, WWP-Itchy and Smurf (with 2–4 WWs; specific to Holozoa). This group also contains two unclassified clades consisting of apusozoan and fungal proteins (with the same protein domain architecture) and a clade with proteins from unicellular Holozoa (with its own domain arrangement consisting of C2 and a CCCH zinc finger). The C2 domain targets the enzyme to membranes by binding to lipids (Ponting and Parker 1996), whereas WW is a recognition domain that selectively picks target proteins, typically through PY motifs (Chen and Sudol 1995; Macias et al. 2002).

A possible explanation for the evolution of this group of families involves the assumption that one ancestral homolog was present in the genome of the last Apusozoa–Opisthokonta common ancestor, which underwent independent diversifications in Apusozoa and Opisthokonta.

The Nedd4 family includes proteins from all holozoan lineages. In animals, Nedd4s are key downregulators of several receptors involved in cell signaling and membrane trafficking. For example, Nedd4s are responsible for the ubiquitination and stability of the insulin-like growth factor I receptor (Vecchione et al. 2003); *Dro. melanogaster* Nedd4 targets Notch receptor for proteasomal degradation (Sakata et al. 2004); and human Nedd4-1 ubiquitinates EGF (epidermal growth factor) receptor and ACK (a tyrosine kinase signaling factor) in response to EGF overexpression itself (Lin et al. 2010).

The WWP-Itchy family is also specific to Holozoa. It includes WWP1, WWP2, and Itchy, three human proteins that have been studied in depth, as well as Su(dx) from *Dro. melanogaster*. WWP-Itchy proteins regulate endosomal sorting and signaling by polyubiquitinating Notch in humans, mice, and *Cae. elegans* (Qiu et al. 2000; Wilkin et al. 2004; Shaye and Greenwald 2005). They also regulate the Hippo pathway: WWP1, WWP2, and Itchy polyubiquitinate AMOT (regulator of YAP/Yorkie, the central member of the Hippo pathway, which is essential for the constitution of a fully functional pathway [Sebé-Pedrós et al. 2012; Wang et al. 2012]). Itchy also polyubiquitinates Warts/Lats, another member of the Hippo pathway found in Opisthokonta (Ho et al. 2011). Moreover, human Itchy polyubiquitinates the transcription factors p63 and p73 (Rossi et al. 2005, 2006).

Within the Smurf family (present in all holozoan lineages except Ichthyosporea), DSmurf (*Dro. melanogaster* homolog) is known to regulate imaginal disc development (Liang et al. 2003) and embryonic dorsal-ventral patterning (Podos et al. 2001) by polyubiquitinating MAD (Dpp pathway); and human Smurfs (Smurf1 and 2) are known to antagonize TGF $\beta$  signaling, and therefore regulate cell growth and proliferation (Massagué and Gomis 2006).

The HECW family (or NEDL/Nedd4-like) contains animal HECTs, including human proteins NEDL1 (which stabilizes p53 in an ubiquitin-independent manner, thereby enhancing p53-mediated apoptosis [Li et al. 2008]) and NEDL2 (which stabilizes p73 [Miyazaki et al. 2003]).

The fungal, apusozoa, and unicellular-holozoan Nedd4-like clades are incertae sedis. As for the Nedd4-like fungal proteins (Fungi clade in fig. 2), only *Saccharomyces cerevisiae* Rsp5p has been characterized: It controls gene expression during nutrient limitation-driven stress (Cardona et al. 2009) and has various roles in intracellular trafficking (Belgareh-Touzé et al. 2008), and plasma membrane and cell wall organization (Kaminska et al. 2005). None of the Nedd4-like proteins from the apusozoa and unicellular-holozoan clade proteins has been characterized.

Class VI also includes several families characterized by a common domain architecture consisting of DUF908, DUF913, and DUF4414 (domains of unknown function). These three domains typically co-occur together in HECT proteins and are evolutionarily conserved in various Unikonta and Bikonta lineages, revealing an ancient origin for this group of proteins. These include HUWE1, HECTFu1 (HUWE1-like), UPL1/2, HECTAI1, and HECTHe3 families.

The HUWE1 family is named after the human protein within it (also known as UREB1, HectH9, KIAA0312, LASU1, ARF-BP1, or Mule). HUWE1 proteins have a complex domain architecture consisting of DUF908, DUF913, WWE, UBA, and DUF4414. It includes representatives from animals, *M. brevicollis* and Amoebozoa. The *M. brevicollis* has a single HECT domain, but proteins from Amoebozoa have the complete arrangement (except WWE). Human HUWE1 polyubiquitinates Myc (oncoprotein and transcription factor), which is essential for the transactivation of several Myc target genes, the recruitment of co-activator p300 and the induction of cell proliferation (Adhikary et al. 2005). It also enhances p53 stability by helping ARF inhibit p53 ubiquitination by Mdm2 (Brooks and Gu 2006), among other functions (Chen et al. 2005; Zhong et al. 2005; Hall et al. 2007).

The HECTFu1 family includes fungal proteins with a HUWE1-like N-terminal architecture (without WWE), and also some specific domains and simpler arrangements. There is indirect evidence that Tom1 (a yeast HUWE1-like protein) intervenes in Cdc6 posttranslational regulation (Hall et al. 2007).

UPL1/2 is a Viridiplantae-specific family that contains Embryophyta proteins with the characteristic DUF908-DUF913-UBA-DUF4414 N-terminal architecture and green algae proteins with a single HECT domain.

Both the HECTHe3 (present in Heterokonta) and HECTAI1 (present in Alveolata) families also contain the DUF4414-HECT arrangement.

Finally, there are four additional families with good nodal support and domain coherence within class VI: KIAA0317, HACE1, HECTHe4, and UPL5.

The HACE1 family contains proteins from all holozoan clades plus *A. castellanii*. HACE1 proteins have a variable number of Ankyrin repeats (typically two to three) and sometimes a PHD domain. The ubiquitinating activity of HACE1 is known to regulate Golgi complex disassembly and reassembly during mitosis (Tang et al. 2011), and also plays a role in various cancer processes (Zhang et al. 2007). The HACE1 and HUWE1 families were thought to be sister groups and, together, to be a sister group to the Nedd4-like group of proteins (Marín 2010); however, we did not recover such topology, but rather a polytomy of several families (fig. 2).

The KIAA0317 family is exclusive to Metazoa and Choanoflagellata (*Sal. rosetta*) clades. Most of them have Filamin repeats, which are only found in this family. They have no

known substrates, but Filamin is known to mediate protein recognition in other proteins and contexts (Ohta et al. 2006).

The HECTHe4 is specific to Heterokonta and includes *P. infestans* proteins with a distinctive zf-RanBP domain and other proteins with a HECT domain. Both ML and BI analyses have linked this family to the Nedd4-like group of proteins, but with low statistical support (fig. 2).

UPL5 is a Bikonta family that includes proteins from Viridiplantae (with a Ub domain), as well as from Rhizaria (*Bigelowiella natans*) and Cryptophyta (*Guillardia theta*) clades (with just a HECT domain). *Arabidopsis thaliana* UPL5 polyubiquitinates the WRKY53 transcription factor, which promotes leaf senescence (Miao and Zentgraf 2010). Ub-like domains within E3 enzymes probably allow for the interaction of these enzymes with other members of the pathway (Miao and Zentgraf 2010).

### The Origins of Multicellularity and the Evolution of the HECT E3 System

As unicellular eukaryotes evolved into multicellular life forms, the need for more complex and finely tuned regulation mechanisms increased and met new regulatory requirements related to cell proliferation, adhesion, differentiation, ordered cell death, and extra/intracellular signaling. Therefore, and given that the ubiquitination pathway is an important regulatory layer responsible for key posttranslational modifications and protein turnover, one may expect expansions of the ubiquitination toolkit (including the HECT system) at the origin of multicellular clades. To ascertain whether this is the case, we analyzed the functional and molecular diversity of the HECT system in several eukaryote lineages.

Specifically, we used the relationship between the number of HECT proteins and the number of distinct N-terminal domain architectures of those proteins as an estimator of the diversity of the HECT system in every given genome. Our data show that the number of HECT proteins positively correlates with the number of distinct N-terminal domain architectures (fig. 6).

According to this, the HECT system is enriched in animals and unicellular Holozoa, the Heterokonta *P. infestans* and *E. siliculosus*, and the Apusozoa *T. trahens*. Conversely, Fungi, plant, Chlorophyta, Rhodophyta, and Excavata genomes are HECT-poor, with fewer proteins and little protein domain diversification. It is worth mentioning that some species such as the Rhizaria *B. natans* and the Haptophyta *Emi. huxleyi* have a high count of HECT proteins but a low degree of domain diversification.

The Apusozoa *T. trahens*, the sister group to Opisthokonta (Torruella et al. 2012), also shows a relatively rich HECT toolkit, much richer than plants and Fungi and similar in complexity to those of metazoans. Our data show that there are some HECT proteins that independently diversified within *T. trahens*. For instance, class I contains an unclassified *T. trahens* clade

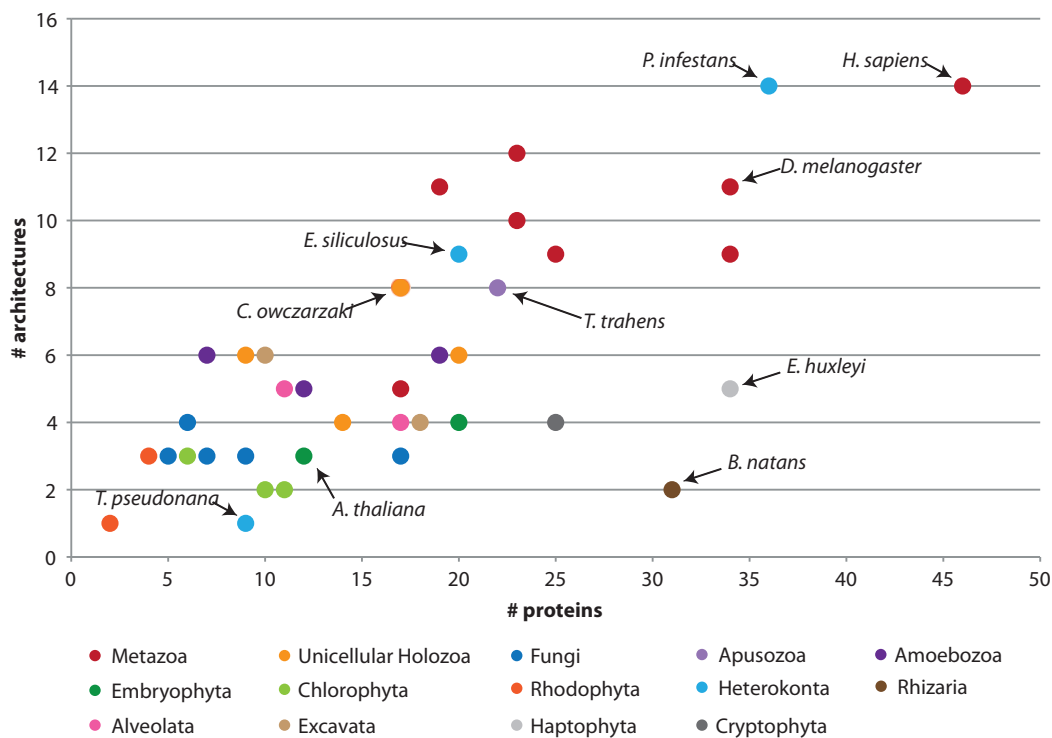
whose proteins have independently acquired different protein recognition domains (such as SPRY, ZZ, and zf-UBR). Also, the well-known Nedd4 group of HECTs dates back to the last common ancestor between Opisthokonta and Apusozoa. New apusozoan genomes will make it possible to gain further insights into the evolution of the HECT system in this lineage.

The diversity of HECTs in Heterokonta is highly variable. *Thalassiosira pseudonana* has a poor HECT system, whereas *E. siliculosus* (a multicellular brown alga) and especially *P. infestans* have a more diversified HECT system comparable with that of animals that most likely evolved from a small basal toolkit similar to that of *Tha. pseudonana*, according to the present phylogeny. Moreover, both *P. infestans* and *E. siliculosus* proteins have convergently acquired several architectures characteristic of Opisthokonta HECTs. For example, *P. infestans* proteins have recognition domains such as MIB-HERC2, UBA, SPRY, or RLD (typical of large HERC families), and *E. siliculosus* proteins have RLD and Kelch repeats.

Our analyses show that animals have the most expanded and diverse HECT system among eukaryotes, and their unicellular holozoan relatives (Choanoflagellata, Filasterea, and Ichthyosporea) have an intermediate diversity of the system (fig. 6). This suggests that there was a burst of HECT diversity at the onset of Metazoa, but that a relatively complex HECT system already existed in the animals' closest unicellular relatives. Indeed, the origin of most (17 out of 22) HECT families containing animal proteins (among those defined in this study) pre-dates the origin of animals (fig. 4). Rather, the higher degree of diversification of HECT in animals is explained by the acquisition of new domains in the N-terminal regions of HECTs. Leaving aside the hemichordate *S. kowalevskii* (a clear outlier to the general trend), animals have between 9 and 14 different HECT architectures, whereas their closest unicellular holozoan relatives have between four and nine arrangements.

The number of families present in each clade provides additional information on the degree of diversification of the HECT system in each taxon (fig. 4). For instance, 24 new families appear at some point during the evolution of the Opisthokonta lineage. The Holozoa are the most family-rich lineage, with 22 families, 5 of which are specific to Metazoa. Also, there are five families present in plants (all of which appear either at the origin of Bikonta or Viridiplantae). This reveals that in both animals and plants most HECT families pre-date the respective origins of multicellularity.

We also mapped the acquisition of N-terminal domains across the tree of eukaryotes (fig. 5). This is a common event within each class, and those architectures that appear at the base of multicellular clades and their closest unicellular relatives are of particular interest. Our data show that the acquisition of new domains is a common event in the holozoan clade, especially at the root of animals and Choanoflagellata (six domains) and at the node leading to Metazoa (eight domains). Indeed, there are five families (namely, EDD, HECTD3, HUWE1, UBE3B, and HERC2) in which animal



**FIG. 6.**—Number of HECT proteins against the number of distinct protein domain architectures found in all HECT-containing proteins for each species. Taxa are color-coded according to taxonomic assignment.

proteins have more complex architectures than those found in their unicellular relatives' homologs. Conversely, the acquisition of specific protein domains in other multicellular lineages such as Fungi and Embryophyta is minimal.

Overall, our data suggest that increases in both N-terminal architectural diversification and absolute number of proteins have shaped the evolutionary history of HECT ligases in eukaryotes. An increase in the protein number brings molecular duplicities that allow sub- or neofunctionalization of HECT proteins. N-terminal domain shuffling is a plastic and adaptable evolutionary mechanism that does not require a change of gene content. It can account for significant evolutionary changes in posttranslational regulation through the adjustment of substrate specificity and protein localization. Indeed, domain shuffling has been acknowledged as an important mechanism for explaining the evolution of multidomain proteins and the appearance of novel proteins, especially regarding the origin of new proteins in major transitions such as the acquisition of multicellularity in animals (Tordai et al. 2005; King et al. 2008; Suga et al. 2012).

It must be noted that HECTs are not the only set of E3 ligases of the ubiquitin system and they are not equally relevant in different eukaryotic lineages. This means that HECT-poor taxa such as plants or Fungi may not necessarily have a poor ubiquitination system. Indeed, *Ara. thaliana*, with just seven HECTs, has expanded their E3 proteins count in terms

of F-box, RING and U-box ligases (Lespinet et al. 2002), compared to other eukaryotes. Conversely, E1 and E2 functions are each performed by a single type of enzymes. All E1 enzymes descend from a common ancestor that was co-opted into ubiquitin activating functions at the origin of eukaryotes, and, since then, has undergone duplications in Unikonta, Vertebrata, Heterokonta, and Kinetoplastida (Excavata) (Burroughs et al. 2009). Similarly, there is just one type of E2 enzyme for conjugating ubiquitin, and all (or most of) their known families were already present at the LECA (Burroughs et al. 2008; Michelle et al. 2009). Altogether, this shows that E1 and E2 enzymes radiated concomitantly prior to the LECA, when they were recruited for the ubiquitination pathway (Burroughs et al. 2008).

This pattern of evolution is markedly different from that showed by HECTs (in this study) and other E3 enzymes (Lespinet et al. 2002), which have undergone differential lineage-specific expansions—in the case of HECTs, those detected in Holozoa, Heterokonta, and maybe Apusozoa. This emphasizes the role of E3s as a specific and functionally specialized step of the ubiquitination pathway.

## Conclusions

Our genomic survey and phylogenetic analysis classifies eukaryotic HECTs in six main classes, whose constituent proteins

probably descend from six ancestral proteins present in the LECA, assuming the “Unikont–Bikont” hypothesis for the rooting of the eukaryote phylogeny. These six classes include 35 identified protein families, as well as other proteins that cannot be classified with certainty.

We also show that, because the eukaryotic ancestor, the HECT system has increased its functional complexity and capacity to finely tune posttranslational protein regulation in several clades, especially—but not exclusively—in multicellular organisms. The system has also been simplified in other clades such as unicellular red algae.

The current diversity of the HECT system has been acquired through two parallel mechanisms: 1) the acquisition of new HECT families through protein duplication, and 2) the acquisition, by domain shuffling, of new protein domains that specifically recognize E3 substrates. We identified a positive correlation between the degree of domain diversification and the number of HECT proteins present in each genome.

Our analysis reveals that this domain syntax of HECT proteins is highly conserved across all eukaryotes: domain fusions always occur at the N-terminus of the proteins. This would be largely due to the physical constraints to catalytic activity imposed by the HECT proteins tertiary structure.

The HECT toolkit evolved in a largely independent manner in different eukaryote clades, often converging in similar domain architectures. Some taxa such as Holozoa are HECT-rich, with many HECT types and various domain arrangements, whereas other taxa such as fungi, plants, and green and red algae have HECT-poor genomes. Regarding the evolution of Holozoa, this study reveals that the onset of new families and new protein recognition motifs typically predate the emergence of animal multicellularity. However, animals further increased their HECT regulatory toolkit from their unicellular ancestor with six new HECT families.

Overall, we show a complex evolutionary scenario in which the HECT system has evolved toward different degrees of diversification in different clades, through family diversification and domain shuffling. Our genomic survey of HECT proteins clarifies the origin and evolution of different HECT families among eukaryotes and also represents a useful evolutionary framework for analyzing this important posttranslational regulatory mechanism.

## Supplementary Material

Supplementary figures S1–S3 and tables S1 and S2 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

The authors thank the Joint Genome Institute and Broad Institute for making data publicly available. This work was supported by an Institució Catalana de Recerca i Estudis

Avançats contract, European Research Council starting grant ERC-2007-StG-206883, Ministerio de Economía y Competitividad (MINECO) grant BFU2011-23434 to I.R.-T., and a pregraduate Formación Profesorado Universitario grant from MINECO to A.S.-P.

## Literature Cited

- Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.
- Adhikary S, et al. 2005. The ubiquitin ligase HectH9 regulates transcriptional activation by Myc and is essential for tumor cell proliferation. *Cell* 123:409–421.
- Aravind L. 2001. The WWE domain: a common interaction module in protein ubiquitination and ADP ribosylation. *Trends Biochem Sci.* 26:273–275.
- Belgareh-Touzé N, et al. 2008. Versatile role of the yeast ubiquitin ligase Rsp5p in intracellular trafficking. *Biochem Soc Trans.* 36:791–796.
- Brooks CL, Gu W. 2006. p53 ubiquitination: Mdm2 and beyond. *Mol Cell.* 21:307–315.
- Brooks WS, et al. 2008. G2E3 is a dual function ubiquitin ligase required for early embryonic development. *J Biol Chem.* 283:22304–22315.
- Burroughs AM, Balaji S, Iyer LM, Aravind L. 2007. Small but versatile: the extraordinary functional and structural diversity of the beta-grasp fold. *Biol Direct.* 2:18.
- Burroughs AM, Jaffee M, Iyer LM, Aravind L. 2008. Anatomy of the E2 ligase fold: implications for enzymology and evolution of ubiquitin/Ubl-like protein conjugation. *J Struct Biol.* 162:205–18.
- Burroughs AM, Iyer LM, Aravind L. 2009. Natural history of the E1-like superfamily: implication for adenylation, sulfur transfer, and ubiquitin conjugation. *Proteins* 75:895–910.
- Cardona F, Aranda A, Del Olmo M. 2009. Ubiquitin ligase Rsp5p is involved in the gene expression changes during nutrient limitation in *Saccharomyces cerevisiae*. *Yeast* 26:1–15.
- Chen D, et al. 2005. ARF-BP1/Mule is a critical mediator of the ARF tumor suppressor. *Cell* 121:1071–1083.
- Chen HI, Sudol M. 1995. The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc Natl Acad Sci U S A.* 92:7819–7823.
- Chong-Kopera H, et al. 2006. TSC1 stabilizes TSC2 by inhibiting the interaction between TSC2 and the HERC1 ubiquitin ligase. *J Biol Chem.* 281:8313–8316.
- Coates J. 2008. Armadillo repeat proteins: versatile regulators of plant development and signalling. *Plant Cell Monogr.* 10:299–314.
- Cruz C, Ventura F, Bartrons R, Rosa JL. 2001. HERC3 binding to and regulation by ubiquitin. *FEBS Lett.* 488:74–80.
- Dastur A, Beaudenon S, Kelley M, Krug RM, Huibregtse JM. 2006. Herc5, an interferon-induced HECT E3 enzyme, is required for conjugation of ISG15 in human cells. *J Biol Chem.* 281:4334–4338.
- Derelle R, Lang BF. 2012. Rooting the eukaryotic tree with mitochondrial and bacterial proteins. *Mol Biol Evol.* 29:1277–1289.
- Downes BP, Stupar RM, Gingerich DJ, Vierstra RD. 2003. The HECT ubiquitin-protein ligase (UPL) family in *Arabidopsis*: UPL3 has a specific role in trichome development. *Plant J.* 35:729–742.
- Gong T-WL, Huang Li, Warner SJ, Lomax MI. 2003. Characterization of the human UBE3B gene: structure, expression, evolution, and alternative splicing. *Genomics* 82:143–152.
- Hadjeji O, Casas-Terradellas E, Garcia-Gonzalo FR, Rosa JL. 2008. The RCC1 superfamily: from genes, to function, to disease. *Biochim Biophys Acta.* 1783:1467–1479.
- Hall J, Kow E, Nevis K, Lu C. 2007. Cdc6 stability is regulated by the Huwe1 ubiquitin ligase after DNA damage. *Mol Biol Cell.* 18:3340–3350.

- Hicke L. 2001. A new ticket for entry into budding vesicles-ubiquitin. *Cell* 106:527–530.
- Ho KC, et al. 2011. Itch E3 ubiquitin ligase regulates large tumor suppressor 1 stability. *Proc Natl Acad Sci U S A.* 108:4870–4875.
- Honda Y, et al. 2002. Cooperation of HECT-domain ubiquitin ligase hHYD and DNA topoisomerase II-binding protein for DNA damage response. *J Biol Chem.* 277:3599–3605.
- Huang L, et al. 1999. Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* 286:1321–1326.
- Itoh M, et al. 2003. Mind bomb is a ubiquitin ligase that is essential for efficient activation of Notch signaling by delta. *Dev Cell.* 4:67–82.
- Jin L, Williamson A, Banerjee S, Philipp I, Rape M. 2008. Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. *Cell* 133:653–665.
- Kajiro M, et al. 2011. The E3 ubiquitin ligase activity of Trip12 is essential for mouse embryogenesis. *PLoS One* 6:e25871.
- Kaminska J, et al. 2005. Rsp5 ubiquitin ligase affects isoprenoid pathway and cell wall organization in *S. cerevisiae*. *Acta Biochim Pol.* 52: 207–220.
- Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33: 511–518.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066.
- Kaustov L, et al. 2007. The conserved CPH domains of Cul7 and PARC are protein-protein interaction modules that bind the tetramerization domain of p53. *J Biol Chem.* 282:11300–11307.
- King N, et al. 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451:783–788.
- Kroismayr R, et al. 2004. HERC5, a HECT E3 ubiquitin ligase tightly regulated in LPS activated endothelial cells. *J Cell Sci.* 117:4749–4756.
- Kühnle S, et al. 2011. Physical and functional interaction of the HECT ubiquitin-protein ligases E6AP and HERC2. *J Biol Chem.* 286: 19410–19416.
- Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286–2288.
- Lee JD. 2002. The ubiquitin ligase hyperplastic discs negatively regulates hedgehog and decapentaplegic expression by independent mechanisms. *Development* 129:5697–5706.
- Lepinet O, Wolf YI, Koonin EV, Aravind L. 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.* 12:1048–1059.
- Liang Y-Y, et al. 2003. dSmurf selectively degrades decapentaplegic-activated MAD, and its overexpression disrupts imaginal disc development. *J Biol Chem.* 278:26307–26310.
- Lin Q, et al. 2010. HECT E3 ubiquitin ligase Nedd4-1 ubiquitinates ACK and regulates epidermal growth factor (EGF)-induced degradation of EGF receptor and ACK. *Mol Cell Biol.* 30:1541–1554.
- Li Y, et al. 2008. A novel HECT-type E3 ubiquitin protein ligase NEDL1 enhances the p53-mediated apoptotic cell death in its catalytic activity-independent manner. *Oncogene* 27:3700–3709.
- Lloyd SE, et al. 2009. HECTD2 is associated with susceptibility to mouse and human prion disease. *PLoS Genet.* 5:e1000383.
- Macias MJ, Wiesner S, Sudol M. 2002. WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett.* 513: 30–37.
- Marín I. 2010. Animal HECT ubiquitin ligases: evolution and functional implications. *BMC Evol Biol.* 10:56–68.
- Maspero E, et al. 2011. Structure of the HECT:ubiquitin complex and its role in ubiquitin chain elongation. *EMBO Rep.* 12:342–349.
- Massagué J, Gomis RR. 2006. The logic of TGFbeta signaling. *FEBS Lett.* 580:2811–2820.
- Miao Y, Zentgraf U. 2010. A HECT E3 ubiquitin ligase negatively regulates *Arabidopsis* leaf senescence through degradation of the transcription factor WRKY53. *Plant J.* 63:179–188.
- Michelle C, Vourc'h P, Mignon L, Andres CR. 2009. What was the set of ubiquitin and ubiquitin-like conjugating enzymes in the eukaryote common ancestor? *J Mol Evol.* 68:616–628.
- Miyazaki K, et al. 2003. A novel HECT-type E3 ubiquitin ligase, NEDL2, stabilizes p73 and enhances its transcriptional activity. *Biochem Biophys Res Commun.* 308:106–113.
- Mukhopadhyay D, Riezman H. 2007. Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science* 315: 201–205.
- Nishiya T, et al. 2011. Regulation of inducible nitric-oxide synthase by the SPRY domain- and SOCS box-containing proteins. *J Biol Chem.* 286: 9009–9019.
- Oeda T, et al. 2001. Oxidative stress causes abnormal accumulation of familial amyotrophic lateral sclerosis-related mutant SOD1 in transgenic *Caenorhabditis elegans*. *Hum Mol Genet.* 10:2013–2023.
- Ohta Y, Hartwig JH, Stossel TP. 2006. FilGAP, a Rho- and ROCK-regulated GAP for Rac binds filamin A to control actin remodelling. *Nat Cell Biol.* 8:803–814.
- Ozols J. 1989. Structure of cytochrome b5 and its topology in the microsomal membrane. *Biochim Biophys Acta.* 997:121–130.
- Pasaje CF, et al. 2011. UBE3C genetic variations as potent markers of nasal polyps in Korean asthma patients. *J Hum Genet.* 56:797–800.
- Peters J-M. 2002. The anaphase-promoting complex: proteolysis in mitosis and beyond. *Mol Cell.* 9:931–943.
- Pickart CM, Fushman D. 2004. Polyubiquitin chains: polymeric protein signals. *Curr Opin Chem Biol.* 8:610–616.
- Podos SD, Hanson KK, Wang YC, Ferguson EL. 2001. The DSmurf ubiquitin-protein ligase restricts BMP signaling spatially and temporally during *Drosophila* embryogenesis. *Dev Cell.* 1:567–578.
- Ponting CP, Parker PJ. 1996. Extending the C2 domain family: C2s in PKCs delta, epsilon, eta, theta, phospholipases, GAPs, and perforin. *Protein Sci.* 5:162–166.
- Punta M, et al. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40:D290–D301.
- Qiu L, et al. 2000. Recognition and ubiquitination of Notch by Itch, a Hect-type E3 ubiquitin ligase. *J Biol Chem.* 275:35734–35737.
- Rhoads A, Friedberg F. 1997. Sequence motifs for calmodulin recognition. *FASEB J.* 11:331–340.
- Rodríguez-Ezpeleta N, et al. 2007. Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Curr Biol.* 17: 1420–1425.
- Rosa J, Casaroli-Marano R. 1996. p619, a giant protein related to the chromosome condensation regulator RCC1, stimulates guanine nucleotide exchange on ARF1 and Rab proteins. *EMBO J.* 15: 4262–4273.
- Rossi M, et al. 2005. The ubiquitin-protein ligase Itch regulates p73 stability. *EMBO J.* 24:836–848.
- Rossi M, et al. 2006. The E3 ubiquitin ligase Itch controls the protein stability of p63. *Proc Natl Acad Sci U S A.* 103:12753–12758.
- Rotin D, Kumar S. 2009. Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol.* 10:398–409.
- Sakata T, Sakaguchi H, Tsuda L, Higashitani A. 2004. *Drosophila* Nedd4 regulates endocytosis of notch and suppresses its ligand-independent activation. *Curr Biol.* 14:2228–2236.
- Sarkar A, Zohn I. 2011. Hectd1 regulates intracellular trafficking of Hsp90 to control its secretion and cell motility of the cranial mesenchyme. *Dev Biol.* 356:120.
- Scheel H, Hofmann K. 2003. No evidence for PHD fingers as ubiquitin ligases. *Trends Cell Biol.* 13:285–7; author reply 287–288.
- Scheffner M. 1998. Ubiquitin, E6-AP, and their role in p53 inactivation. *Pharmacol Ther.* 78:129–139.

- Schwartz AL, Ciechanover A. 2009. Targeting proteins for destruction by the ubiquitin system: implications for human pathobiology. *Annu Rev Pharmacol Toxicol.* 49:73–96.
- Sebé-Pedrós A, Zheng Y, Ruiz-Trillo I, Pan D. 2012. Premetazoan origin of the hippo signaling pathway. *Cell Rep.* 1:13–20.
- Shaye DD, Greenwald I. 2005. LIN-12/Notch trafficking and regulation of DSL ligand activity during vulval induction in *Caenorhabditis elegans*. *Development* 132:5081–5092.
- Smits V. 2012. EDD induces cell cycle arrest by increasing p53 levels. *Cell Cycle* 11:715–720.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Stechmann A, Cavalier-Smith T. 2002. Rooting the eukaryote tree by using a derived gene fusion. *Science* 297:89–91.
- Suga H, et al. 2012. Genomic survey of premetazoans shows deep conservation of cytoplasmic tyrosine kinases and multiple radiations of receptor tyrosine kinases. *Sci Signal.* 5:ra35.
- Tang D, et al. 2011. The ubiquitin ligase HACE1 regulates Golgi membrane dynamics during the cell cycle. *Nat Commun.* 2:501.
- Tewari R, Bailes E, Bunting KA, Coates JC. 2010. Armadillo-repeat protein functions: questions for little creatures. *Trends Cell Biol.* 20:470–481.
- Tomaic V, et al. 2011. Regulation of the human papillomavirus type 18 E6/E6AP ubiquitin ligase complex by the HECT domain-containing protein EDD. *J Virol.* 85:3120–3127.
- Tordai H, Nagy A, Farkas K, Bányai L, Patthy L. 2005. Modules, multidomain proteins and organismic complexity. *FEBS J.* 272:5064–5078.
- Torruella G, et al. 2012. Phylogenetic relationships within the opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Mol Biol Evol.* 29:531–544.
- Vecchione A, Marchese A, Henry P. 2003. The Grb10/Nedd4 complex regulates ligand-induced ubiquitination and stability of the insulin-like growth factor I receptor. *Mol Cell Biol.* 23:3363–3372.
- Verdecia MA, et al. 2003. Conformational flexibility underlies ubiquitin ligation mediated by the WWP1 HECT domain E3 ligase. *Mol Cell.* 11:249–259.
- Wang C, et al. 2012. The Nedd4-like ubiquitin E3 ligases target angiomin/p130 to ubiquitin-dependent degradation. *Biochem J.* 444:279–289.
- Wilkin MMB, et al. 2004. Regulation of notch endosomal sorting and signaling by *Drosophila* Nedd4 family proteins. *Curr Biol.* 14:2237–2244.
- Wu W, et al. 2010. HERC2 is an E3 ligase that targets BRCA1 for degradation. *Cancer Res.* 70:6384–6392.
- Ying M, Zhan Z, Wang W, Chen D. 2009. Origin and evolution of ubiquitin-conjugating enzymes from *Guillardia theta* nucleomorph to hominoid. *Gene* 447:72–85.
- Yu J, et al. 2008. The E3 ubiquitin ligase HECTD3 regulates ubiquitination and degradation of Tara. *Biochem Biophys Res Commun.* 367:805–812.
- Zhang L, et al. 2007. The E3 ligase HACE1 is a critical chromosome 6q21 tumor suppressor involved in multiple cancers. *Nat Med.* 13:1060–1069.
- Zhang L, Kang L, Bond W, Zhang N. 2009. Interaction between syntaxin 8 and HECTd3, a HECT domain ligase. *Cell Mol Neurobiol.* 29:115–121.
- Zhong Q, Gao W, Du F, Wang X. 2005. Mule/ARF-BP1, a BH3-only E3 ubiquitin ligase, catalyzes the polyubiquitination of Mcl-1 and regulates apoptosis. *Cell* 121:1085–1095.

Associate editor: Purificación López-García