Polyglutamine expansion in *Drosophila:* thermal stress and Hsp70 as selective agents

BRIAN R BETTENCOURT*, CATHERINE C HOGAN and MARIO NIMALI
Department of Biological Sciences, University of Massachusetts Lowell, 1 University Ave., Lowell, MA 01854, USA
**Corresponding author (Fax, 1-978-934 3044; Email, Brian_Bettencourt@uml.edu)*

Repetitive DNA sequences that encode polyglutamine tracts are prone to expansion and cause highly deleterious phenotypes of neurodegeneration. Despite this tendency, polyglutamine tracts ("polyQs") are conserved features of eukaryotic genomes. PolyQs are the most frequent protein-coding homotypic repeat in insect genomes, and are found predominantly in genes encoding transcription factors conserved from *Drosophila* through human. Although highly conserved across species, polyQ lengths vary widely within species. In *D. melanogaster*, polyQs in 25 genes have more alleles and higher heterozygosity than all other poly-amino acid tracts. The heat shock protein Hsp70 is a principal suppressor of polyQ expansions and may play a key role in modulating the phenotypes of the alleles that encode them. Hsp70 also promotes tolerance of natural thermal stress in *Drosophila* and diverse organisms, a role which may deplete the chaperone from buffering against polyQ toxicity. Thus in stressful environments, natural selection against long polyQ alleles more prone to expansion and deleterious phenotypes may be more effective. This hypothesis can be tested by measuring the phenotypic interactions between *Hsp70* and polyQ transgenes in *D. melanogaster* undergoing natural thermal stress, an approach which integrates comparative genomics with experimental and ecological genetics.

[Bettencourt B R, Hogan C C and Nimali M 2007 Polyglutamine expansion in *Drosophila*: thermal stress and Hsp70 as selective agents; *J. Biosci.* **32** 537–547]

1. Introduction

Stress is as ancient as life itself, and shapes the evolution of genes, genomes, and organisms – a fact attested to by the kingdom-spanning conservation of heat shock proteins (Hsps) and Hsp70 in particular. Hsp70 is the principal eukaryotic stress-inducible heat shock protein and promotes the ability of organisms to survive multiple environmental stresses. Since its first discovery as a heat-induced puff on *Drosophila* polytene chromosomes (Ritossa 1962), Hsp70 has starred in an exhaustive list of studies linking its expression to environmental stress tolerance in virtually every organism examined, and research in the field is seemingly mature.

Recently however, biologists investigating phenomena otherwise unrelated to environmental stress uncovered a genetic stress linked to Hsp70: polyglutamine expansion (Sakahira *et al* 2002). This highly deleterious stress originates within the cell, is encoded by organisms' own genomes, and is as ancient, conserved, and thus fundamental to Hsp70's evolution as external environmental stresses (*see* Alba *et al* 2001 and below). This paper will examine seemingly juxtaposed conservation and variability of polyglutamine-encoding genes, review the links between polyglutamine expansion and Hsp70, and suggest a role for environmental stress in modulating Hsp70's effect on polyglutamine evolution. Finally, we outline an approach to rigorously test the ideas discussed within that integrates comparative genomics, experimental molecular biology, and population genetics.

2. Polyglutamine expansion – a stress encoded by eukaryotic genomes

Repetitive DNA sequences are ubiquitous components of most genomes, especially in eukaryotes. Simple di- and trinucleotide

Keywords. CAG repeats; heat shock proteins; Hsp70; Polyglutamine

repeats represent a distinct subtype of DNA repeat especially problematic for the cell's replication machinery. These repeats are prone to expansion due to polymerase slippage, caused by secondary structure formation within the repeats themselves (Pearson *et al* 2005). Within protein-coding genes, repeats of the glutamine-encoding CAG and CAA codons are particularly problematic. Slippage on CAG/CAA repeats is frequent, and allelic diversity of length variants is very high (Leggo *et al* 1997; Kittles *et al* 2001; Fujii *et al* 2003). These "polyglutamine tracts" (polyQs) can dramatically modify the phenotypes of their host proteins: long polyQ tracts cause deleterious phenotypes in many organisms. In humans, expanded polyQs cause at least nine severe human neurodegenerative disorders, including Huntington's and Kennedy's diseases, several spinocerebellar ataxias (SCAs), and dentatorubralpallidoluysian atrophy (DRPLA; reviewed in Everett and Wood 2004). Disease phenotypes are generally caused by dominant alleles encoding polyQ tracts longer than ~37 amino acid residues (Schaffar *et al* 2004). Expanded polyQ tracts self-aggregate into fibrils, wherein the polyQs form stable "polar zippers" of parallel β sheets (Ross *et al* 1998). Importantly, expanded polyQs can also coaggregate with other endogenous proteins that contain shorter, nonpathogenic polyQ repeats (Gatchel and Zoghbi 2005). As polyQ domains are found in numerous (non-disease) proteins including crucial transcription factors such as TATA-binding protein and the coactivator CREB-binding protein, coaggregation with expanded polyQs and loss of function by sequestration of the aggregates is a critical aspect of polyQ toxicity (Wyttenbach 2004). Whether cytotoxic polyQ aggregation itself, and/or depletion of endogenous proteins via sequestration, are the root of pathogenesis remains an area of open investigation (Herbst and Wanker 2006). Both types of aggregation link polyQ expansion to molecular chaperones such as Hsp70 (*see* below).

2.1 *PolyQs are highly conserved in transcription factors*

Given their deleterious phenotypes, evolution would seem to favour elimination of polyQ tracts. However, polyQs are present and highly conserved in diverse eukaryotic genomes including those of primates, other mammals, insects, and yeast (Zoghbi and Botas 2002; Driscoll and Gerstbrein 2003). Many polyQ tracts are functional components of conserved *trans*-activation domains in transcription factors (Callewaert *et al* 2003; Huang *et al* 2003; Ambra *et al* 2004; Sheu *et al* 2006), and disruption of polyQ-encoding transcription factor

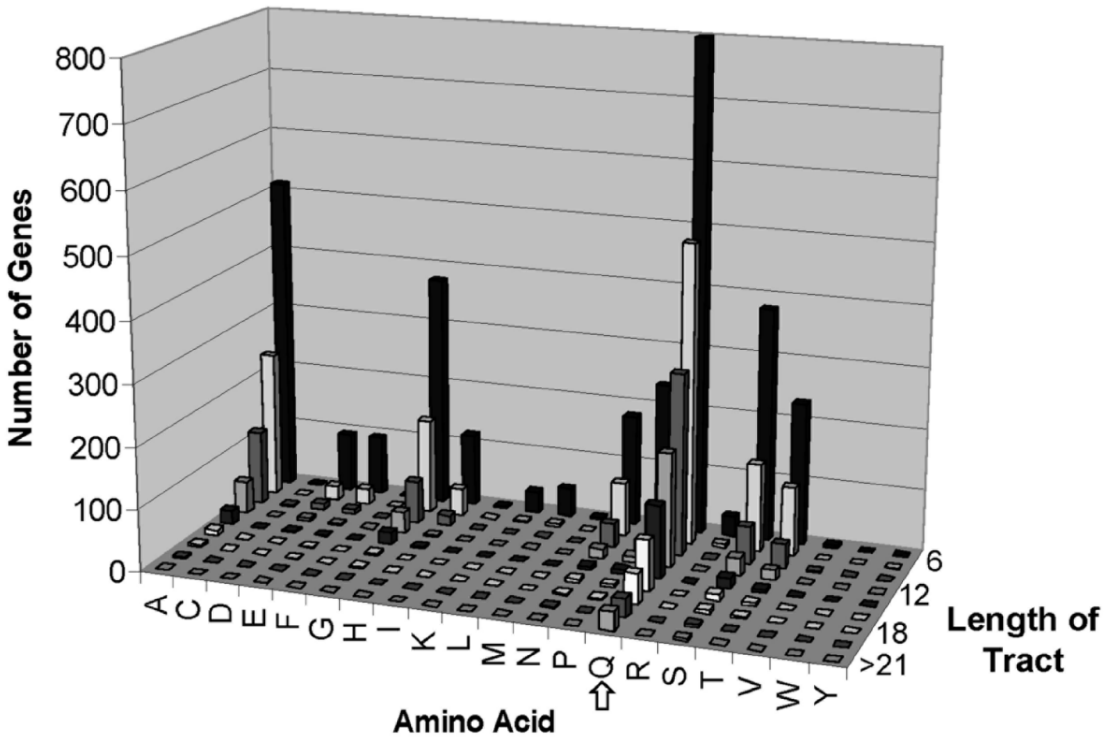

Figure 1. Length distribution of homotypic poly-amino acid tracts in the *D. melanogaster* Release 3.1 protein-coding genome. One transcript per gene was downloaded from *www.fruitfly.org*, translated, and scanned for uninterrupted repeats of greater than five residues using custom perl software (available upon request). Arrow marks polyglutamine (Q), the most abundant repeat.

activity is the critical lesion in many polyQ diseases (Riley and Orr 2006). Similarly, many polyQ-containing disease genes are conserved across eukaryotes, and a number of nonhuman models of polyQ diseases have been recently advanced (e.g. mouse, Mitsui *et al* 2002; Sakai *et al* 2006; *Drosophila*, Kazemi-Esfarjani and Benzer 2000, Chan *et al* 2002; nematode, Brignull *et al* 2006; yeast, Krobitsch and Lindquist 2000). In the animal models, human polyQ proteins effectively elicit the same host responses as endogenous polyQs, including aggregation, toxicity, and

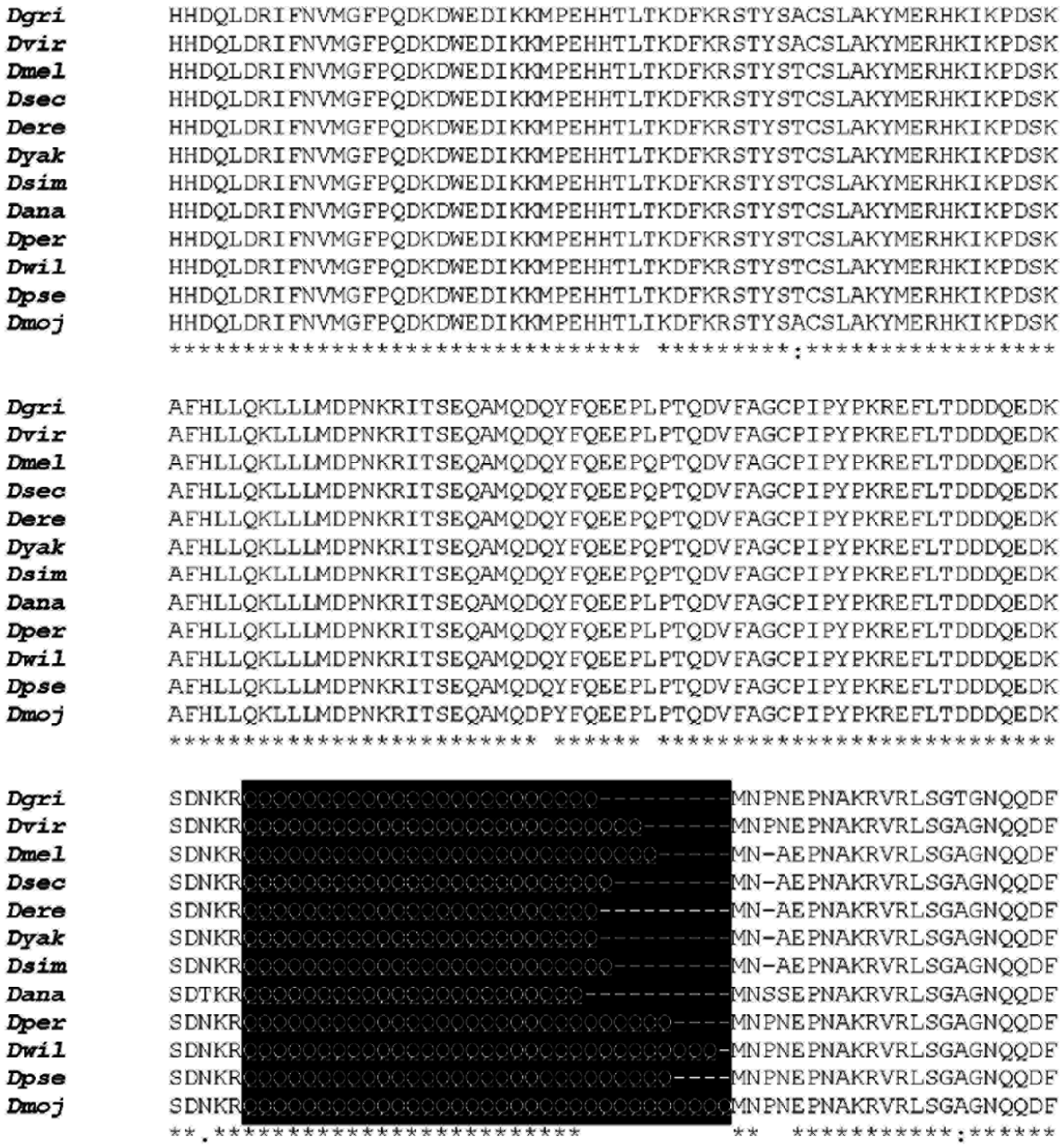

Figure 2. Partial protein sequence alignment of the *D. melanogaster (Dmel)* transcriptional mediator Cyclin-Dependent Kinase 8 (CG10572) and GENIE-predicted orthologs from other *Drosophila* species. Three-letter species abbreviations as found at *http://rana.lbl.gov/drosophila*. Polyglutamine tracts highlighted in black. Stars, colons, and periods mark identical residues, conservative replacements, and semiconservative replacements, respectively.

Table 1. Twenty-five *D. melanogaster* genes conserved in all twelve *Drosophila* species (fourth column) possessing conserved polyglutamine tracts of twenty residues or greater.

Gene	GO Term(s)	Human ortholog	12 *Dros*	*Agam*
CG2904 (echinus)	Thiolesterase	ubiquitin-specific protease 53	Yes	Yes
CG4013 (Smr)	Transcription repressor	nuclear receptor corepressor 1	Yes	Yes
CG11172 (NFAT)	Transcription factor	nuclear factor of activated T cells 5	Yes	Yes
CG12236	Transcription factor	–	Yes	Yes
CG12653 (buttonhead)	Transcription factor	Sp1 transcription factor	Yes	No
CG12690 (CHES-like)	Transcription factor	checkpoint suppressor 1	Yes	Yes
CG32532	Transcription factor	aristaless-related homeobox	Yes	Yes
CG32548		-	Yes	No
CG32778	Transcription factor	myelin transcription factor-like 1	Yes	No
CG1775 (medea)	DNA binding	mothers against DPP homolog 4	Yes	Yes
CG3978 (pannier)	Transcription factor	GATA-binding protein	Yes	No
CG5466		–	Yes	No
CG12071	DNA binding	–	Yes	No
CG17228 (prospero)	Transcription factor	prospero-related homeobox 1	Yes	Yes
CG32466 (rotund)	Transcription factor	nuclear matrix transcription factor 4	Yes	Yes
CG33106 (mask)	DNA binding	CXXC finger 6	Yes	Yes
CG2083		ankyrin-repeat domain protein	Yes	Yes
CG6619		GRP-associated scaffold protein	Yes	Yes
CG7391 (Clock)	Transcription factor	Clock	Yes	Yes
CG10572 (Cdk8)	Transcription factor	Cdk8	Yes	Yes
CG3695 (Trap150β)	Transcription mediator	CSRP3	Yes	Yes
CG4744		–	Yes	Yes
CG5580 (scribbler)	Transcription factor	zinc finger protein 608	Yes	No
CG10543	Transcription factor	zinc finger protein 567	Yes	Yes
CG14023		ALR	Yes	Yes

GO Term(s), annotations from FlyBase (*www.flybase.net*). Human ortholog, annotation of *Homo sapiens* gene identified by BLASTP homology with expectation value less than E^{-10}, *Agam,* similar identification of ortholog in the mosquito *Anopheles gambiae*.

heat shock protein expression. PolyQ *function* is conserved across broad phylogenetic ranges. PolyQ *sequences* may, therefore, be subject to opposing selective pressures: selection for maintenance of their function *vs.* selection against their toxic expansion-prone tendencies.

The recent proliferation of sequenced insect genomes that complement the well-annotated *D. melanogaster* genome offers a great opportunity to examine polyQ evolution. In the *D. melanogaster* protein-coding genome, polyQs are the most frequent type of homotypic repeat greater than 5 residues long (figure 1). There are at least 296 *D. melanogaster* genes encoding 10 or more residues long polyQs, and 30 genes encode 20 or more residues long polyQs (figure 1). What are these genes?

The genes in *D. melanogaster* that encode long polyQs are largely transcription factors. Twenty-two of 30 are annotated as such in FlyBase, a significantly higher fraction than is found among functional annotations in the entire genome (22/30 *vs.* 992/13604; Chi-squared test, $P < 0.001$). This finding is consistent with the observation of polyQ domains in other eukaryotic transcription factors (*see* above). These proteins are highly conserved across 12 *Drosophila* species genomes with, on average, greater than 75% amino acid identity (unpublished data; for example *see* figure 2). Twenty-six (of 30) are found via 1:1 reciprocal best BLAST hit clustering in all 12 species (unpublished; for explanation of methods *see* Richards *et al* 2005). For 25 of these ortholog sets, the polyQ tracts are conserved in all *Drosophila* species, with average polyQ length conservation exceeding 50% (table 1). Similarly, orthologs can be identified for 20 of the polyQ proteins in the mosquito *Anopheles gambiae*, and 20 in humans (table 1). Thus polyQ

Table 2. Heterozygosity and number of alleles at polyglutamine-encoding (left) and other poly-amino acid-encoding (right) genes in a Pennsylvanian (USA) population of *D. melanogaster*. Heterozygosity calculated as one minus the sum of the squared allele frequencies. "Repeat AA" indicates repeated amino acid residue. Bottom right: Summary statistics and unpaired *T*-tests comparing polyglutamine (PolyQ) genes with other amino acid repeat genes (PolyOther).

Gene	Repeat AA	Heterozygosity	Alleles
CG2904(echinus)	Q	0.71	5
CG4013(634->827)	Q	0.49	2
CG4013(973->1124)	Q	0.61	4
CG11172 (NFAT)	Q	0.42	2
CG12236	Q	0.28	2
CG12653 (buttonhead)	Q	0.49	3
CG12690 (CHES)	Q	0.82	8
CG32532	Q	0.32	2
CG32548	Q	0.81	6
CG32778	Q	0.76	5
CG1775 (medea)	Q	0.3	2
CG3978 (pannier)	Q	0.55	4
CG5466	Q	0.15	2
CG12071	Q	0.32	2
CG17228 (prospero)	Q	0.75	4
CG32466 (rotund)	Q	0.78	5
CG2083	Q	0.37	2
CG6619	Q	0.62	3
CG7391 (Clock)	Q	0.76	5
CG10572 (Cdk8)	Q	0.61	3
CG3695	Q	0.46	2
CG4744	Q	0.83	7
CG5580 (scribbler)	Q	0.71	4
CG10543	Q	0.5	2
CG14023	Q	0.28	2
CG1133 (OPA)	Q	0.15	2
CG1210 (PK61C)	Q	0.51	3
CG1322 (Zfh1)	Q	0	1
CG1343 (Sp1)	Q	0.7	4
CG1379	Q	0.61	3
CG1693 (tty)	Q	0.49	2
CG9755 (pumilio)	Q	0.74	5
CG9731	Q	0.6	3
CG9369 (miniature)	Q	0.47	3

Gene	Repeat AA	Heterozygosity	Alleles
CG10777	A	0	1
CG11245	A	0	1
CG11486	A	0.28	2
CG13188	A	0.5	2
CG33106 (mask)	A	0.15	2
CG3373 (Hemomucin)	T	0.53	2
CG15046	T	0.3	2
CG11073	T	0	1
CG12586	T	0.5	2
CG13194	T	0	1
CG7139	N	0	1
CG33175 (Sprint)	N	0.44	2
CG32790	N	0.69	4
CG12212 (pebbled)	N	0.7	4
CG15455	N	0.28	2
CG4341	S	0.15	2
CG6066	S	0.5	2
CG14478	S	0.5	2
CG31475	S	0.69	4
CG15784	S	0.48	2
CG13316 (Mnt)	S	0.72	2
CG8108	D	0.45	2
CG8184	D	0	1
CG8545	D	0.65	4

	PolyQ	PolyOther	*P* value
Mean Heterozygosity:	0.5285294	0.366	0.01
Mean Number of Alleles:	3.3529412	2.12	0.001

orthologs are found in 66-87% of comparisons (depending on species examined), a fact that speaks of their high degree of sequence conservation and lack of paralogs. By comparison, for all 13456 *D. melanogaster* protein-coding genes, 12-*Drosophila* species ortholog clusters are identified for only 4164, or 31%, using identical methods (data not shown). This very broad conservation of polyQ protein sequence and function as transcription factors suggests that polyQs persist – despite their known proclivity toward problems – due to selection for function.

2.2 *Although highly conserved across species, polyQs vary within species*

Despite the high conservation among polyQ-encoding genes between species, the lengths of polyQ tracts vary within species. CAG/CAA repeats, often studied as anonymous microsatellite loci, display extreme length variability in organisms as diverse as flies, plants, and worms (Michalakis and Veuille 1996; Huttel *et al* 1999; Haber *et al* 2005). Within known protein-coding genes, where the repeats definitely encode polyQs, variability is similarly high (Deka *et al* 1999; Andres *et al* 2003; Sheu *et al* 2006). In a single natural population of *D. melanogaster*, 25 genes encoding polyQs of 20 or more residues long have significantly more length-variant alleles and higher heterozygosity than all other similarly sized poly-amino acid tracts in the genome (table 2). Given the high GC content of CAG/CAA repeats, their variability might simply be a reflection of high secondary structure stability and thus frequent replicative slippage. However, many of the other poly-amino acid repeats in *D. melanogaster* are encoded by equally GC-rich sequences, and none display the length variability or heterozygosity of the polyQs [e.g. polyD (GAC) or polyA (GCA) repeats, *see* table 2]. This indicates that while other poly-amino acid repeats may slip and expand, the resulting alleles are likely neutral ones whose frequencies are affected by drift. In contrast, the polyQs have more alleles at higher frequencies – indicating that their frequencies may be affected by selection. Thus, the polyQs are distinctive in terms of within-species variability, are extremely conserved across species, and possess potent phenotypes. The preservation of polyQs across evolutionary time is thus not likely to be a static phenomenon. Among mechanisms that may modulate polyQ evolution, Hsp70 is a particularly attractive candidate.

3. HSP70 suppresses polyQ expansion and phenotypes

In diverse organisms, the Hsps, and Hsp70 in particular, function as effective suppressors of polyQ expansion and toxicity (Wyttenbach *et al* 2002; Ishihara *et al* 2003; Smith *et al* 2005). Genetic screens in *Drosophila* and other organisms have identified Hsp70 and its cofactor Hsp40 as the most potent suppressors of polyQ toxicity (Warrick *et al* 1999; Fernandez-Funez *et al* 2000; Kazemi-Esfarjani and Benzer 2000, 2002). Hsp70 can bind expanded polyQ tracts and prevent aggregation (Kim *et al* 2004; Wacker *et al* 2004), and Hsp overexpression can suppress polyQ disease phenotypes in many systems, including *Drosophila* (reviewed in Muchowski and Wacker 2005, *see* also Katsuno *et al* 2005). By binding and sequestering expanded polyQs quickly after they are expressed, Hsp70 can also prevent the cascade of expanded polyQ/native protein interactions that may be pathogenic. Thus, Hsp70 can provide a buffer against the constant input of expressed expanded polyQs (Wyttenbach 2004). Hsp70's role in responding to polyQ expansion is consistent with a model of polyQ toxicity termed "chaperone depletion": Expanded polyQs can bind chaperones such as Hsp70, deplete the cell of chaperone activity necessary for many critical protein folding/function tasks, and thus cause deleterious phenotypes (Bonini 2002).

3.1 *Stress from within vs. stress from without – a zero-sum game?*

Hsp70 must also respond to acute environmental stresses such as heat shock. Hsp70 is a primary molecular chaperone that protects cells from thermally-induced protein misfolding (Young *et al* 2004). During heat shock, Hsp70 binds denatured proteins and either assists their refolding or sequesters them to prevent cytotoxic aggregate formation. Hsp70 expression is tightly regulated and highly inducible, and promotes fitness via tolerance of naturally occurring thermal stress in numerous systems including *Drosophila* (Kregel 2002). Hsp70 expression in the absence of stress, however, is deleterious (Feder *et al* 1992; Krebs and Feder 1997); thus selection for thermotolerance is opposed by selection against overexpression. Hsp70's high degree of response to thermal stress raises a challenge to buffering of polyQ toxicity. Unlike acute external stresses such as heat shock, input of expanded polyQs into the cell is a constant related to the expression level of the polyQ-encoding gene. Thus, the requirement for Hsp70's buffering of polyQs will persist. If Hsp70 is recruited to cope with heat shock, will its polyQ-suppressive phenotype be reduced? Conceptually, this "environmental chaperone depletion" model is the logical converse of the chaperone depletion model of polyQ toxicity (above), and suggests that depletion of Hsp70's activity by thermally denatured proteins will render cells deficient in buffering against expanded polyQ aggregation and toxicity.

3.2 *The model: environmental stress, Hsp70 expression and selection on polyQ expansion*

If cells are unable to buffer polyQ phenotypes when Hsp70 is depleted due to thermal stress, then natural selection against longer polyQ alleles throughout the genome may become more effective: When stress is frequent, long polyQs will be more likely to cause deleterious phenotypes and thus be culled from the population. When stress is less frequent, Hsp70 will be more able to suppress polyQ phenotypes and thus render polyQ length variation phenotypically "cryptic" (*sensu* Rutherford and Lindquist 1998). These observations

predict an inverse relationship between the frequency of thermal stress and both the average length and diversity of polyQ alleles in natural populations. Expansion produces longer polyQ alleles, selection favors shorter ones – and the interaction of Hsp70 expression with environmental stress determines where "balance" may lie. Testing whether, how, and why this phenomenon operates in nature necessitates a diverse set of experiments, and thus demands an integrated approach. Such an approach must minimally incorporate comparative genomics, experimental molecular biology, and population genetics. *D. melanogaster* offers the ideal blend of genomic resources, molecular genetic tractability, and ecology to serve as a model system for this integrative research.

4. Testing the model

An integrated research program employing *D. melanogaster* to examine the relationship between environmental stress, Hsp70, and polyQ evolution involves progress in at least three of the following areas.

4.1 *First: Identify the genes encoding polyQs, and determine whether they are targets of natural selection*

The *D. melanogaster* protein-coding genome is superbly annotated based on multiple types of evidence (Misra *et al* 2002). Thus in general, the search for polyQ genes is easily accomplished via PERL or similar programming languages, and turns up well-supported, accurate gene models. Finding orthologs in other genomes involves generating and parsing large amounts of pairwise blast data, then computing clusters of orthologous genes – an analysis rendered routine by the common availability of computing clusters and distributed grid systems. This paper does not detail methods for ortholog identification and analysis (pending publication of the eleven *Drosophila* species genomes; however, *see* Richards *et al* 2005 for essentials). However, it may be noted that since putative polyQ-encoding tracts are repetitive, many gene prediction/finding programs skip them, and some human annotators may hesitate to assign them as coding regions. As such, manual examination of unfiltered translated blast data is often necessary. Once orthologs are identified and aligned, the overall degree of polyQ sequence conservation can be measured, albeit often after manual inspection and editing of gaps imposed by software-specific handling of repetitive sequences. Given a well-aligned set of orthologous polyQ genes, mapping which lineage(s) and which gene(s) experience polyQ expansion and contraction is possible, but not demonstrative of selection *per se*. Instead, gene-centric computational methods that hunt for the molecular signatures of recent natural selection must be employed on the polyQ genes. For example, traditional categorical tests or frequency-spectrum analyses of sequence divergence and polymorphism (*e.g.* McDonald-Kreitman tests, K_A / K_S, Tajima's D) can indicate when polyQ genes are evolving under purifying or natural selection. However, as these types of tests focus on *the gene* as a whole rather than the polyQ region specifically, they may be inadequate or misleading. Codon-based approaches that measure selection at particular residues represent an improvement (Yang and Nielsen 2002), but still measure the likelihood of natural selection based on models of nonrepetitive protein sequence evolution that may not apply to polyQs. A candidate alternative approach, specific to polyQ-encoding loci, will examine what proportion of each polyQ is encoded by pure CAG *vs*. mixed CAG/CAA repeats. Pure CAG repeats form stronger hairpins supported by more C:G base pairs than mixed repeats, and may be more prone to slippage, expansion, and phenotypic consequence (Sobczak and Krzyzosiak 2005). Alleles composed primarily of CAG repeats are thus likely young, non-neutral, and subject to natural selection. Mixed CAG/CAA repeats on the other hand may be more stable in length, older, and a result of purifying selection maintaining a base level of polyQ length. We are currently comparing polyQ lengths and CAG/CAA content across *Drosophila* species and within *D. melanogaster* populations, to determine which if any polyQ genes and repeats are subject to natural selection. Initial research indicates that within populations, polyQ tracts are extremely variable in length, despite a high degree of cross-species conservation (*see* above). Key questions currently under investigation include: Are longer polyQ alleles higher in CAG content than shorter alleles? Are longer alleles more recently derived, and at high or low frequency in populations? Are mixed CAG/CAA repeat regions subject to purifying selection and preserved across species? Finally, do pure CAG repeats display increased divergence between species?

Examining polyQ gene sequences to measure whether/when/how selection acts on them is a first step, and one that will require further conceptual and statistical advances to characterize neutral and non-neutral patterns of repetitive sequence evolution. The next critical step is to determine whether and how Hsp70 interacts with polyQs to alter their exposure to natural selection.

4.2 *Second: Test whether variation in Hsp70 expression during naturally occurring thermal stress modulates the phenotypes of specific polyQ alleles*

PolyQ lengths vary substantially in *D. melanogaster* – but how is Hsp70 specifically responsible for shaping this

variation, if at all? To address this question, experiments must conclusively link variation in Hsp70 expression to variation in polyQ phenotypes consequential for fitness. The goal is to simultaneously assay the phenotypes of both specific polyQ tracts and specific Hsp70 mutations by exploiting the transgenic tractability of *D. melanogaster*.

Nancy Bonini and coworkers recently established a *D. melanogaster* model of human polyglutamine diseases by creating a fly strain that expresses a 78-residue polyQ, derived from the human Machado-Joseph disease protein, in the developing eye (Bonini 2002). The "Q78" strain displays marked eye degeneration, with easily measurable phenotypes including defects in pigmentation, facet irregularity, and necrosis (Bonini 2002; *see* figure 3). To investigate whether altered Hsp70 expression would affect the eye phenotypes, we crossed Q78 flies with "TraIII" flies bearing six additional *Hsp70* genes (twice the wildtype *Hsp70* genomic copy number, *see* Welte *et al* 1993). Similarly, we crossed Q78 flies with flies lacking half or all *Hsp70* genes ("*Hsp70A*$^{-}$*Ba*$^{-}$" and "*Hsp70*$^{-}$", *see* Gong and Golic 2006). The resulting panel of lines shares a common genetic background, expresses polyQ with discernable phenotypes, and differs only by specific *Hsp70* mutations that alter Hsp70 expression. Q78 flies wildtype for *Hsp70* display reduced pigment deposition and moderate irregularity (figure 3C), while Q78; *Hsp70*$^{-}$ flies have severely deformed eyes that are almost devoid of pigment, irregular, and marked by necrotic patches (figure 3A). This is especially striking since all the Q78; Hsp70 mutant strains possess additional miniwhite transgenes when compared to the Q78 strain and thus can potentially express more red pigment (for example, compare figure 3C, D). Q78; *Hsp70A*$^{-}$*Ba*$^{-}$ flies are intermediate in appearance (figure 3B). In contrast, Q78; TraIII flies are almost entirely wildtype in appearance, indicating that Hsp70 can suppress the polyQ eye phenotype (figure 3D). PolyQ expression in the Q78 line is driven by the GAL4/UAS system and is thus sensitive to varying rearing temperatures (N Bonini, pers. comm.). Satisfyingly, the effect of rearing temperature mimics that of *Hsp70* genotype in the combination Q78; Hsp70 mutant lines. For example, rearing Q78 flies at 26°C yields morphologically irregular eye phenotypes resembling those of Q78; *Hsp70*$^{-}$ flies reared at 22°C (figure 3E), albeit with less pigmentation due to differences in transgenic construction (*see* above). Similarly, Q78; *Hsp70*$^{-}$ flies reared at 26°C are extremely irregular in eye morphology (figure 3F) and often die prematurely (data not shown). Finally, rearing Q78; *Hsp70*$^{-}$ flies at 18°C produces eyes wildtype in morphology and almost wildtype in pigmentation (intermediate between eyes illustrated in figure 3C, D).

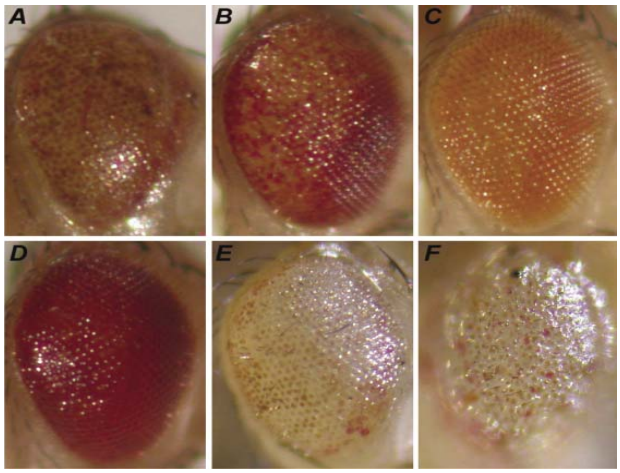

Figure 3. Eyes of one-day-old male *D. melanogaster* of the following strains and rearing temperatures: (A) Q78; *Hsp70*$^{-}$ at 22°C, (B) Q78; *Hsp70A*$^{-}$*Ba*$^{-}$ at 22°C, (C) Q78 at 22°C, (D) Q78; TraIII at 22°C, (E) Q78 at 26°C, and (F) Q78; *Hsp70*$^{-}$ at 26°C. All strains possess one polyQ construct-bearing second chromosome, one balancer second chromosome (*GMR-GAL4, UAS-Q78(S) / SM6A*), and are homozygous for their respective third chromosomes.

How can the Q78; Hsp70 mutant lines be used to test whether Hsp70 expression modifies polyQ phenotypes in nature? The next critical steps will be: first, heat shock the lines in the laboratory to measure the effects of Hsp70 expression at defined temperatures on the severity of the polyQ eye phenotypes. This will determine whether acute thermal stress depletes or enhances Hsp70 suppression of polyQs. Second, extend the laboratory experiments to natural *Drosophila* habitats, where thermal stress varies according to environmental factors such as solar exposure.

Feder and coworkers successfully employed this approach – "seeding" apples with larvae of different *Hsp70* genotypes, monitoring the apples' thermal profiles *in situ,* recovering larvae for Hsp70 expression analysis, and counting survivors – to link Hsp70 expression to survival of naturally occurring heat stress (Feder *et al* 1996; Roberts and Feder 2000). Similar experiments with the Q78; Hsp70 mutant lines will establish whether natural stress induction of Hsp70 is sufficient to affect polyQ suppression (as measured by reduced numbers of flies with deformed eyes and/or reduction in the severity of deformity). Roberts and Feder (2000) established that in apples exposed to full sun, TraIII larval thermotolerance is significantly higher than wild-type; will the Q78; TraIII line also suppress polyQ phenotypes to a greater degree than wildtype? In the shade, the Q78; TraIII line may display even greater suppression of polyQs, since there is no other challenge for the extra Hsp70 expression provided by the *Hsp70* transgenes.

The *Hsp70*$^{-}$ line is severely deficient in inducible thermotolerance (Gong and Golic 2006). Thus the Q78; *Hsp70*$^{-}$ line should be both less thermotolerant and less able to suppress polyQ phenotypes in the sun. Critically, the Q78; *Hsp70*$^{-}$ line ought to show greatly reduced polyQ

suppression in the shade, since it can not express Hsp70 in response to either heat or polyQ expression. The polyQ and thermotolerance phenotypes of the Q78; *Hsp70A*$^{-}$*Ba*$^{-}$ line may be intermediate, depending on the severity of thermal stress and the degree to which half the wildtype copy number of *Hsp70* genes is sufficient to promote its survival. If successful, the above experiments will demonstrate how polyQ suppression will be altered when *transgenes* manipulate the response of Hsp70 to natural thermal stress. The final necessary step will be similar demonstration when *natural selection* manipulates Hsp70 expression.

4.3 *Third: Measure the relationship between polyQ allele frequencies and Hsp70 expression in populations inhabiting natural stress gradients*

To determine whether and how Hsp70 modulation of polyQ expansion affects polyQ alleles throughout the genome, the links between Hsp70, polyQ phenotypes, and stress can be extended to populations inhabiting a natural stress gradient. Populations inhabiting the eastern coast of Australia are an ideal candidate. These populations experience variation in frequency of natural thermal stress that is correlated with latitude (Bettencourt *et al* 2002). This variation in stress drives natural selection on variation in thermotolerance and *Hsp70* alleles that differ in their Hsp70 expressivity (Bettencourt *et al* 2002; Weeks *et al* 2002). In the Australian populations, Hsp70 expression may thus also vary with latitude. If this clinal variation drives differential suppression of (and thus selection on) polyQ phenotypes, will a consistent relationship between latitude and polyQ lengths result? This question can be addressed by measuring polyQ length distribution at multiple genes in the Australian populations, and comparing allele frequencies with those at other poly-amino acid repeat genes (*see* above) and additional neutral length-variable markers. The relationship between polyQ lengths and latitude may not be a simple linear one, as is seen with *Hsp70* allele frequencies. A consistent relationship between polyQ alleles and latitude – coupled with the lack of such a relationship for other poly-amino acid genes and neutral markers – would suggest a causal link between stress, Hsp70 expression and polyQs in natural populations.

5. Conclusions

Polyglutamine expansion is a genetic stress encoded by diverse eukaryotic genomes and conserved in the sequences of transcription factors, whose deleterious phenotypes are buffered by the heat shock protein Hsp70. Despite the extremely broad conservation of the genes encoding polyQs, polyQ tracts vary extensively within natural populations, suggesting frequent natural selection on their lengths. Environmental stress may be an important selective agent modulating polyQ lengths, as mediated through Hsp70 expression. Determining whether and how stress, natural selection, and the interaction of Hsp70 with polyQ expansion shapes the evolution of polyQ sequences throughout the genome will require experimental advances on a number of fronts and thus necessitates an integrated approach. Recent progress in understanding the basic biology of polyQs and Hsp70, especially in the context of human disease, has been brisk. Leveraging this progress into similar advances in our understanding of how polyQs and Hsp70 may co-evolve is a significant challenge – one worth undertaking as the insights provided may aid in our basic understanding of why such a deleterious phenotype persists in species and populations.

Acknowledgements

We would like to thank Subhash Lakhotia and Martin Feder for organizing the Symposium on Environmental Factors, Cellular Stress and Evolution, the INDO-US Forum for providing travel support, and the US Department of Defense for research funding.

References

Alba M M, Santibanez-Koref M F and Hancock J M 2001 The comparative genomics of polyglutamine repeats: extreme differences in the codon organization of repeat-encoding regions between mammals and *Drosophila*; *J. Mol. Evol.* **52** 249–259

Ambra R, Grimaldi B, Zamboni S, Filetici P, Macino G and Ballario P 2004 Photomorphogenesis in the hypogeous fungus *Tuber borchii*: isolation and characterization of Tbwc-1, the homologue of the blue-light photoreceptor of *Neurospora crassa*; *Fungal Genet. Biol.* **41** 688–697

Andres A M, Lao O, Soldevila M, Calafell F and Bertranpetit J 2003 Dynamics of CAG repeat loci revealed by the analysis of their variability; *Hum. Mutat.* **21** 61–70

Bettencourt B R, Kim I, Hoffmann A A and Feder M E 2002 Response to natural and laboratory selection at the *Drosophila Hsp70* genes; *Evolution* **56** 1796–1801

Bonini N M 2002 Chaperoning brain degeneration; *Proc. Natl. Acad. Sci. USA* **99 S4** 16407–16411

Brignull H R, Moore F E, Tang S J and Morimoto R I 2006 Polyglutamine proteins at the pathogenic threshold display neuron-specific aggregation in a pan-neuronal *Caenorhabditis elegans* model; *J. Neurosci.* **26** 7597–7606

Callewaert L, Christiaens V, Haelens A, Verrijdt G, Verhoeven G and Claessens F 2003 Implications of a polyglutamine tract in the function of the human androgen receptor; *Biochem. Biophys. Res. Commun.* **306** 46–52

Chan E Y, Luthi-Carter R, Strand A, Solano S M, Hanson S A, DeJohn M M, *et al* 2002 Increased huntingtin protein length reduces the number of polyglutamine-induced gene expression changes in mouse models of Huntington's disease; *Hum. Mol. Genet.* **11** 1939–1951

Deka R, Guangyun S, Smelser D, Zhong Y, Kimmel M and Chakraborty R 1999 Rate and directionality of mutations and effects of allele size constraints at anonymous, gene-associated, and disease-causing trinucleotide loci; *Mol. Biol. Evol.* **16** 1166–1177

Driscoll M and Gerstbrein B 2003 Dying for a cause: invertebrate genetics takes on human neurodegeneration; *Nat. Rev. Genet.* **4** 181–194

Everett C M and Wood N W 2004 Trinucleotide repeats and neurodegenerative disease; *Brain* **127** 2385–2405

Feder J H, Rossi J M, Solomon J, Solomon N and Lindquist S 1992 The consequences of expressing *Hsp70* in *Drosophila* cells at normal temperatures; *Genes Dev.* **6** 1402–1413

Feder M E, Blair N and Figueras H 1996 Natural thermal stress and heat-shock protein expression in *Drosophila* larvae and pupae*; Func. Ecol.* **11** 90–100

Fernandez-Funez P, Nino-Rosales M L, de Gouyon B, She W C, Luchak J M, Martinez P, *et al* 2000 Identification of genes that modify ataxin-1-induced neurodegeneration; *Nature (London)* **408** 101–106

Fujii K, Okiura T, Nishimura K, Orimoto C and Nishimukai H 2003 Short tandem repeat typing in exon 1 of the androgen receptor gene; *Leg. Med. (Tokyo)* **5** S201–S203

Gatchel J R and Zoghbi H Y 2005 Diseases of unstable repeat expansion: mechanisms and common principles; *Nat. Rev. Genet.* **6** 743–755

Gong W J and Golic K G 2006 Loss of Hsp70 in *Drosophila* is pleiotropic, with effects on thermotolerance, recovery from heat shock and neurodegeneration; *Genetics* **172** 275–286

Haber M, Schungel M, Putz A, Muller S, Hasert B and Schulenburg H 2005 Evolutionary history of *Caenorhabditis elegans* inferred from microsatellites: evidence for spatial and temporal genetic differentiation and the occurrence of outbreeding; *Mol. Biol. Evol.* **22** 160–173

Herbst M and Wanker E E 2006 Therapeutic approaches to polyglutamine diseases: combating protein misfolding and aggregation; *Curr. Pharm. Des.* **12** 2543–2555

Huang H, Rambaldi I, Daniels E and Featherstone M 2003 Expression of the Wdr9 gene and protein products during mouse development; *Dev. Dyn.* **227** 608–614

Huttel B, Winter P, Weising K, Choumane W, Weigand F and Kahl G 1999 Sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.); *Genome* **42** 210–217

Ishihara K, Yamagishi N, Saito Y, Adachi H, Kobayashi Y, Sobue G, *et al* 2003 Hsp105alpha suppresses the aggregation of truncated androgen receptor with expanded CAG repeats and cell toxicity; *J. Biol. Chem.* **278** 25143–25150

Katsuno M, Sang C, Adachi H, Minamiyama M, Waza M, Tanaka F, *et al* 2005 Pharmacological induction of heat-shock proteins alleviates polyglutamine-mediated motor neuron disease; *Proc. Natl. Acad. Sci. USA* **102** 16801–16806

Kazemi-Esfarjani P and Benzer S 2000 Genetic suppression of polyglutamine toxicity in *Drosophila*; *Science* **287** 1837–1840

Kazemi-Esfarjani P and Benzer S 2002 Suppression of polyglutamine toxicity by a *Drosophila* homolog of myeloid leukemia factor 1; *Hum. Mol. Genet.* **11** 2657–2672

Kim W Y, Horbinski C, Sigurdson W and Higgins D 2004 Proteasome inhibitors suppress formation of polyglutamine-induced nuclear inclusions in cultured postmitotic neurons; *J. Neurochem.* **91** 1044–1056

Kittles R A, Young D, Weinrich S, Hudson J, Argyropoulos G, Ukoli F, *et al* 2001 Extent of linkage disequilibrium between the androgen receptor gene CAG and GGC repeats in human populations: implications for prostate cancer risk; *Hum. Genet.* **109** 253–261

Kregel K C 2002 Heat shock proteins: modifying factors in physiological stress responses and acquired thermotolerance; *J. Appl. Physiol.* **92** 2177–2186

Krebs R A and Feder M E 1997 Deleterious consequences of Hsp70 overexpression in *Drosophila melanogaster* larvae; *Cell Stress Chap.* **2** 60–71

Krobitsch S and Lindquist S 2000 Aggregation of huntingtin in yeast varies with the length of the polyglutamine expansion and the expression of chaperone proteins; *Proc. Natl. Acad. Sci. USA* **97** 1589–1594

Leggo J, Dalton A, Morrison P J, Dodge A, Connarty M, Kotze M J, *et al*. 1997 Analysis of spinocerebellar ataxia types 1, 2, 3, and 6, dentatorubral-pallidoluysian atrophy, and Friedreich's ataxia genes in spinocerebellar ataxia patients in the UK; *J. Med. Genet.* **34** 982–985

Michalakis Y and Veuille M 1996 Length variation of CAG/CAA trinucleotide repeats in natural populations of Drosophila melanogaster and its relation to the recombination rate; *Genetics* **143** 1713–1725

Misra S, Crosby M A, Mungall C J, Matthews B B, Campbell K S, Hradecky P, *et al* 2002 Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review; *Genome Biol.* **3** RESEARCH0083

Mitsui K, Nakayama H, Akagi T, Nekooki M, Ohyawa K, Takio K, *et al* 2002 Purification of polyglutamine aggregates and identification of elongation factor-1a and heat shock protein 84 as aggregate-interacting proteins; *J. Neurosci.* **22** 9267–9277

Muchowski PJ and Wacker JL 2005 Modulation of neurodegenera-tion by molecular chaperones; *Nat. Rev. Neurosci.* **6** 11–22

Pearson C E, Edamura K N and Cleary J D 2005 Repeat instability: mechanisms of dynamic mutations; *Nat. Rev. Genet.* **6** 729–742

Richards S, Liu Y, Bettencourt B R, Hradecky P, Letovsky S, Nielsen R, *et al* 2005 Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution; *Genome Res.* **15** 1–18

Riley B E and Orr H T 2006 Polyglutamine neurodegenerative diseases and regulation of transcription: assembling the puzzle; *Genes Dev.* **20** 2183–2192

Ritossa F 1962 A new puffing pattern induced by temperature shock and DNP in *Drosophila*; *Experentia* **18** 571–573

Roberts S P and Feder M E 2000 Changing fitness consequences of *Hsp70* copy number in transgenic *Drosophila* undergoing natural thermal stress; *Func. Ecol.* **14** 353–357

Ross C A, Margolis R L, Becher M W, Wood J D, Engelender S, Cooper J K, *et al*. 1998 Pathogenesis of neurodegenerative diseases associated with expanded glutamine repeats: new answers, new questions; *Prog. Brain. Res.* **117** 397–419

Rutherford S L and Lindquist S 1998 Hsp90 as a capacitor for morphological evolution; *Nature. (London)* **396** 336–342

Sakahira H, Breuer P, Hayer-Hartl M K and Hartl F U 2002 Molecular chaperones as modulators of polyglutamine protein aggregation and toxicity; *Proc. Natl. Acad. Sci. USA (Suppl 4.)* **99** 16412–16418

Sakai K, Yamada M, Sato T, Yamada M, Tsuji S and Takahashi H 2006 Neuronal atrophy and synaptic alteration in a mouse model of dentatorubral-pallidoluysian atrophy; *Brain* **129** 2353–2362

Schaffar G, Breuer P, Boteva R, Behrends C, Tzvetkov N, Strippel N, *et al*. 2004 Cellular toxicity of polyglutamine expansion proteins: mechanism of transcription factor deactivation; *Mol. Cell.* **15** 95–105

Sheu Y T, Zmuda J M, Cauley J A, Moffett S P, Rosen C J, Ishwad C, *et al* 2006 Nuclear receptor coactivator-3 alleles are associated with serum bioavailable testosterone, insulin-like growth factor-1, and vertebral bone mass in men; *J. Clin. Endocrin. Metab.* **91** 307–331

Smith R C, Rosen K M, Pola R and Magrane J 2005 Stress proteins in Alzheimer's disease; *Int. J. Hyperthermia* **21** 421–431

Sobczak K and Krzyzosiak W J 2005 CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts; *J. Biol. Chem.* **280** 3898–3910

Wacker J L, Zareie M H, Fong H, Sarikaya M and Muchowski P J 2004 Hsp70 and Hsp40 attenuate formation of spherical and annular polyglutamine oligomers by partitioning monomer; *Nat. Struct. Mol. Biol.* **11** 1215–1222

Warrick J M, Chan H Y E, Gray-Board G L, Chai Y H, Paulson H L and Bonini N M 1999 Suppression of polyglutamine-mediated neurodegeneration in *Drosophila* by the molecular chaperone HSP70; *Nat. Genet.* **23** 425–428

Weeks A R, McKechnie S W and Hoffmann A A 2002 Dissecting adaptive clinal variation: markers, inversions and size/stress associations in *Drosophila melanogaster* from a central field population; *Ecol. Lett*. **5** 756–763

Welte M A, Tetrault J M, Dellavalle R P and Lindquist S L 1993 A new method for manipulating transgenes: engineering heat tolerance in a complex, multicellular organism; *Curr. Biol.* **3** 842–853

Wyttenbach A, Sauvageot O, Carmichael J, Diaz-Latoud C, Arrigo A P and Rubinsztein D C 2002 Heat shock protein 27 prevents cellular polyglutamine toxicity and suppresses the increase of reactive oxygen species caused by huntingtin; *Hum. Mol. Genet.* **11** 1137–1151

Wyttenbach A 2004 Role of heat shock proteins during polyglutamine neurodegeneration; *J. Mol. Neurosci.* **23** 69–96

Yang Z and Nielsen R 2002 Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages; *Mol. Biol. Evol.* **19** 908–917

Young J C, Agashe V R, Siegers K and Hartl F U 2004 Pathways of chaperone-mediated protein folding in the cytosol; *Nat. Rev. Mol. Cell. Biol.* **5** 781–791

Zoghbi H Y and Botas J 2002 Mouse and fly models of neurodegeneration; *Trends Genet.* **18** 463–471

*e*Publication: 15 March 2007