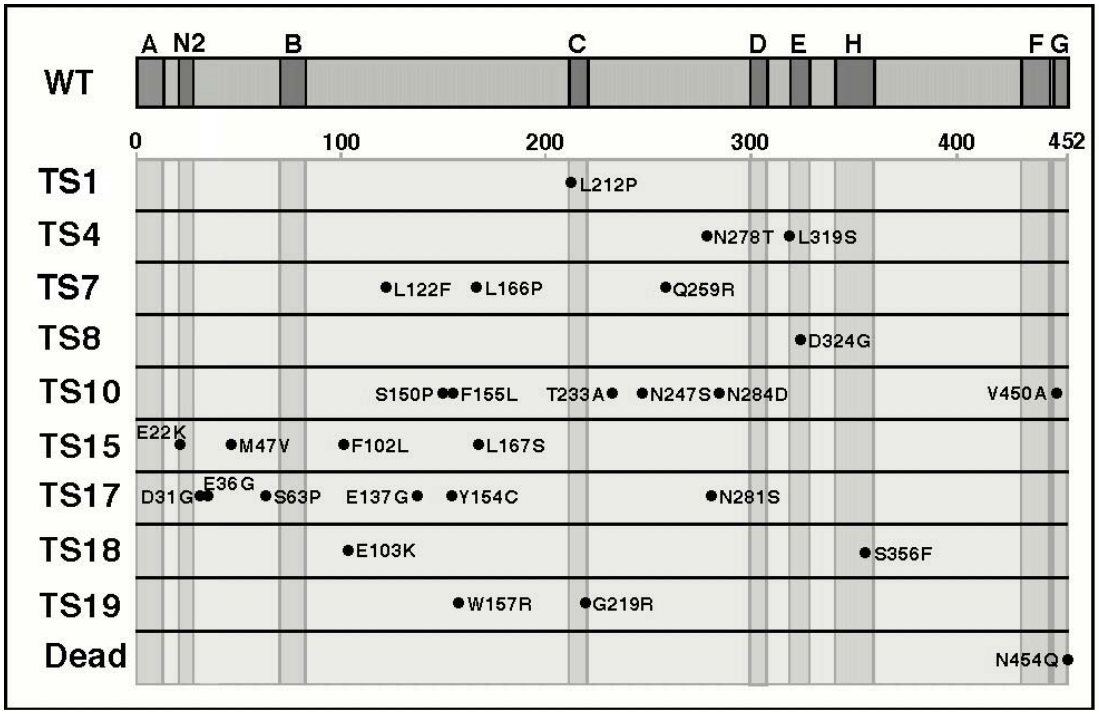
Supplementary Figure 1 Mutations in intein^{TS} alleles.



A schematic representation of a wild type VMA1 intein highlighting the regions (A-H & N2) conserved between inteins and indicating amino acid positions (Top). See www.neb.com/inteins/intein_intro.html and [1] regarding conserved regions. (Bottom) The position and amino-acid substitution caused by the missense mutations identified in the TS alleles isolated. Only mutations resulting in amino acid substitutions are indicated using the single letter code.

The nine TS alleles identified were sequenced to identify the lesions responsible for their phenotype. While no mutations were found in the DNA encoding the flanking regions of the Gal4 host protein, intein sequence analysed indicated a mutation rate of 0.3% per base. Of the total changes, 10 conservative and 27 missense mutations were identified while nonsense and frame shift mutations were not recovered. The positions of the mutations identified with relation to the structure of the intein and its conserved domains are shown. No attempt has been made to identify the individual contributions of each lesion within multiply mutated intein molecules although comparison to the TS mutations recently described in the *Mycobacterium xenopi* Gyrase A intein may prove to be informative[2]. It should also be noted that the full length Sce VMA intein used contains the SceI restriction endonuclease in addition to the domains required for splicing[3]. The SceI region includes the C, D, E and H domains and is not required for protein splicing

activity[3, 4]. While it is likely that the mutations identified in these domains do not affect splicing directly, it remains possible that they may influence protein stability.

1. Pietrokovski, S. Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci.* **7**, 64-71 (1998).
2. Adam, E. & Perler, F.B. Development of a positive genetic selection system for inhibition of protein splicing using mycobacterial inteins in Escherichia coli DNA gyrase subunit A. *Mol Microbiol Biotechnol.* **4**, 479-487 (2002).
3. Chong, S. & Xu, M.Q. Protein splicing of the Saccharomyces cerevisiae VMA intein without the endonuclease motifs. *J Biol Chem.* **272**, 15587-15590 (1997).
4. Noren, C.J., Wang, J. & Perler, F.B. Dissecting the Chemistry of Protein Splicing and Its Applications. *Angew Chem Int Ed Engl* **39**, 450-466 (2000).

Supplementary Table 1 Insertion context of inteinTS alleles.

Postion	Sequence	Comments
VMA1	DAIIYVG/**C...N**/CGE	endogenous context
Gal4 C21	CRLKKLK/**C...N**/CSK	
Gal80 C127	FVEWALA/**C...N**/CSL	did not work in *Drosophila*
Gal80 C277	NGNVPVS/**C...N**/CSF	

The context of the Sce VMA intein in its endogenous host protein (VMA1) and the locations used in this study. The ends of intein sequences are shown in **BOLD** while host sequences are shown as NORMAL text.

The host residues immediately N-terminal to an intein have been described as potentially playing a role in determining splicing activity. Although these residues do not play a direct role in the splicing reaction, their carbonyl carbon atoms are the target of three nucleophilic displacements during the course of the auto-catalytic splicing event[1]. However, while *in vitro* studies suggest that residues most similar to the cognate residue in this position (a Glycine residue in VMA1, see table above) are most effective[2] this appears to be less critical *in vivo* as the insertions generated for this project follow Lysine, Alanine and Serine residues, all of which were previously predicted to splice poorly[2]. Therefore, despite our successful generation of Gal4INT and Gal80INT alleles, this may not be mirrored in all host protein contexts[3] and more widespread application of this technique will be required before the facility of the system can be fully assessed. A more extensive analysis comparing the different flanking sequences may give additional insight into the context requirements.

1. Noren, C.J., Wang, J. & Perler, F.B. Dissecting the Chemistry of Protein Splicing and Its Applications. *Angew Chem Int Ed Engl* **39**, 450-466 (2000).
2. Chong, S. et al. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucleic Acids Res* **26**, 5109-5115 (1998).
3. Wu, W., Wood, D.W., Belfort, G., Derbyshire, V. & Belfort, M. Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing. *Nucleic Acids Res* **30**, 4864-4871 (2002).

Supplementary Table 2 Oligonucleotides used.

Name	Sequence (5' - 3')	Comments
G4D02	CAAAATCATGTCAAGGTC	Gal4 sequencing
G80DP1R	GGTTACCAGATCTACACCGTTCCCGATTTCATAGATGTATTT TGGTGACTGATGGACGACAACCTGGTTG	INTEIN^{Dead} amplification for insertion into Position 1
G80DP2R	AATTTCTGCGAATCCGGCATCGCCTTCAAGTTTCAAATCTCC CTTGGTCTGATGGACGACAACCTGGTTG	INTEIN^{Dead} amplification for insertion into Position 2
G80DP3R	ACGTTCAGCAGCAGCCTTATAAATGGATTCGGCTTGATCTAG TGAACACTGATGGACGACAACCTGGTTG	INTEIN^{Dead} amplification for insertion into Position 3
G80DP4R	CAAATTTTTGGTAAATTTTTTGGTAGGTTTGCCACCTTTGAA ACTGCACTGATGGACGACAACCTGGTTG	INTEIN^{Dead} amplification for insertion into Position 4
G80P1Fc	AGATTGCTGGAAATGGCGGTTGGTACGGCTACGAAAGGCCT GTTAAATGCTTTGCCAAGGGAACGAATGT	INTEIN amplification for insertion into Position 2
G80P1Rb	GGTTACCAGATCTACACCGTTCCCGATTTCATAGATGTATTT TGGTGAATTATGGACGACAACCTGGTTG	INTEIN amplification for insertion into Position 2
G80P2Fc	AATTTACCAAAAAATTTACCAAAAATTTGGTCATTGACATTC ACGGTTGCTTTGCCAAGGGAACGAATGT	INTEIN amplification for insertion into Position 1
G80P2Rb	AATTTCTGCGAATCCGGCATCGCCTTCAAGTTTCAAATCTCC CTTGGTATTATGGACGACAACCTGGTTG	INTEIN amplification for insertion into Position 1
G80P3F	AAAATAATCCGAACCTCAAGTATCTTTTCGTAGAATGGGCCC TTGCATGCTTTGCCAAGGGAACGAATGT	INTEIN amplification for insertion into Position 3
G80P3R	ACGTTCAGCAGCAGCCTTATAAATGGATTCGGCTTGATCTAG TGAACAATTATGGACGACAACCTGGTTG	INTEIN amplification for insertion into Position 3
G80P4F	ATCTTTTATTCCAAGGCACATTGTTAAATGGCAATGTTCCAG TGTCATGCTTTGCCAAGGGAACGAATGT	INTEIN amplification for insertion into Position 4
G80P4R	CAAATTTTTGGTAAATTTTTTGGTAGGTTTGCCACCTTTGAA ACTGCAATTATGGACGACAACCTGGTTG	INTEIN amplification for insertion into Position 4
Gal401-F	CGGGATCCCTACGTAATGCACGCCATC	amplification of Gal4 and addition of BamHI site
Gal402-R	GGGGTACCTACAGATCCTCCTCGGAGATCAGCTTCTGCTCTT TTTTTGGGTTTGGTGG	amplification of Gal4 and addition of Myc tag and EroRI site
Gal405-R	CTCCCAGTTGTTCTTAAGACACTTGGCGCACTTCGGTTTTTCT TTGGAGCACTTAAGCTTTTTAAG	introduces silent HindIII and AflII mutations into Gal4
Gal406c-F	CCCAAGCTTAAATGCTTTGCCAAGGGAACGAATGTTTTAATG	adds HindIII site and some Gal4 sequence to 5' end of INTEIN
Gal407-R	AGTCCACTCAAAATAGGCTTTATTTGAAGC	internal INTEIN primer to destroy endogenous HindIII site
Gal408-R	GTTGTTCTTAAGACACTTGGCGCACTTCGGTTTTTCTTTGGA GCAATTATGGACGACAACCTGG	adds AflII site and some Gal4 sequence to 3' end of INTEIN
Gal80F2	ATAGGATCCTTGCGGCCGCAACATGGACTACAAC	amplification of Gal80
Gal80R3	AAGAGCTCAATCTAGATTATAAACTATAATGC	amplification of Gal80
INS1	ATGAAGCTACTGTCTTCTATC	INTEIN sequencing
INS2	CTTTCTTTTGGCCCATCTC	INTEIN sequencing
INS3B	GTAATGCGACCCATGAGTTG	INTEIN sequencing
INS4	AGAAGTATCTCTGGAATC	INTEIN sequencing
INS5	ACTTGCTTATTTACTTGG	INTEIN sequencing
INS6	GACATATAAATAGCATAC	INTEIN sequencing
INS7	TGCTGGTCTAATTGATTC	INTEIN sequencing
INT01-F	GTCGGGTGCTTTGCCAAGGGTACC	amplification of full length INTEIN
INT02-R	CTCCGCAATTATGGACGACAACCTGG	amplification of full length INTEIN
INT05-F	CAACCAGGTTGTCGTCCATCAGTGCTCCAAAGAAAAACCG	mutates INTEIN N645Q to give dead INTEIN
P1F	TCACCAAAATACATCTATGA	linearisation of pS5-Gal80 at position 1

P1R	TTTAACAGGCCTTTCGTAGC	linearisation of pS5-Gal80 at position 1
P3F	TGTTCTAGATCAAGCCGAATC	linearisation of pS5-Gal80 at position 3
P3R	TGCAAGGGCCCATTCTACGAAAAG	linearisation of pS5-Gal80 at position 3
P4F	TGCAGTTTCAAAGGTGGCAAACC	linearisation of pS5-Gal80 at position 4
P4R	TGACACTGGAACATTGCCATTTAAC	linearisation of pS5-Gal80 at position 4
TIF	TTGCTTGCCAACCAGGTTGTC	INTEIN internal primer to test for insertion in GAL80
TP1F	GGGTTCAAACCATCATCTC	test for insertion in Gal80 P1
TP1R	GATCCGGTACTGTCTTTGGG	test for insertion in Gal80 P1
TP2F	GATTGGGCCAGCGAGTCCCA	test for insertion in Gal80 P2
TP2R	ATGAATGTTACCCACAATGGG	test for insertion in Gal80 P2
TP3F	CCCACTTTAGAGTCATTTG	test for insertion in Gal80 P3
TP3R	GATTTAACAGGCCTTTCG	test for insertion in Gal80 P3
TP4F	CAACATTTGGTCACACAATC	test for insertion in Gal80 P4
dppNotI	ATAACTATGCGGCCGCTTTGTTCAATTTTGTAACAG	cloning *dpp-hsp70* promoter
hsp70SpeI	AGACTAGTTGTGTGTGAGTTCTTCTTTC	cloning *dpp-hsp70* promoter

The oligonucleotides used during the course of this project are described in the text. In general PCR primer pairs carry either the R or F suffix to indicate reverse or forward primers. All sequences are shown 5' to 3'.