The FTO Gene, Implicated in Human Obesity, Is Found Only in Vertebrates and Marine Algae

Steven Robbens, Pierre Rouzé, J. Mark Cock, Jürg Spring, Alexandra Z Worden, Yves van de Peer

J Mol Evol
DOI 10.1007/s00239-007-9059-z

LETTER TO THE EDITOR

The *FTO* Gene, Implicated in Human Obesity, Is Found Only in Vertebrates and Marine Algae

Steven Robbens · Pierre Rouzé · J. Mark Cock · Jürg Spring · Alexandra Z. Worden · Yves Van de Peer

Received: 16 October 2007 / Accepted: 7 November 2007

Abstract Human obesity is a main cause of morbidity and mortality. Recently, several studies have demonstrated an association between the *FTO* gene locus and early onset and severe obesity. To date, the *FTO* gene has only been discovered in vertebrates. We identified *FTO* homologs in the complete genome sequences of various evolutionary diverse marine eukaryotic algae, ranging from unicellular photosynthetic picoplankton to a multicellular seaweed. However, *FTO* homologs appear to be absent from all other completely sequenced genomes of plants, fungi, and invertebrate animals. Although the biological roles of these marine algal *FTO* homologs are still unknown, these genes will be useful for exploring basic protein features and could hence help unravel the function of the *FTO* gene in vertebrates and its inferred link with obesity in humans.

Electronic Supplementary Material The online version of this article (doi:10.1007/s00239-007-9059-z) contains supplementary material, which is available to authorized users.

S. Robbens · P. Rouzé · Y. Van de Peer (✉)
Department of Plant Systems Biology, VIB, B-9052 Ghent, Belgium
e-mail: yves.vandepeer@psb.ugent.be

S. Robbens · Y. Van de Peer
Department of Molecular Genetics, Ghent University, B-9052 Ghent, Belgium

P. Rouzé
Laboratoire Associé de l'INRA (France), Ghent University, B-9052 Ghent, Belgium

J. M. Cock
UMR 7139 CNRS-UPMC, Végétaux Marins et Biomolécules, Station Biologique, BP74, 29682 Roscoff, France

J. Spring
Institute of Zoology, University of Basel, CH-4056 Basel, Switzerland

A. Z. Worden (✉)
Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039-9644, USA
e-mail: azworden@mbari.org

Obesity is a major societal issue contributing to increased morbidity and mortality, as well as rising health care costs. In 2003–2004, 66% of the human population in the United States was classified as overweight (body mass index [BMI] $\geq$ 25 kg/m^2), and 32% was classified as obese (BMI $\geq$ 30 kg/m^2) (Ogden et al. 2006). Excessive weight is often associated with an increased risk of several life-threatening diseases, including cancer, heart diseases, and type 2 diabetes mellitus (Frayling et al. 2007). Unfortunately, the number of obese people continues to increase every day, probably as a result of a modified lifestyle (more food and less exercise). An improved understanding of the genetic basis, and the associated risk factors, is necessary if society is to proactively address this epidemic. Recently, several studies have demonstrated an association between the *FTO* gene locus and early onset and severe obesity in both children and adults (Dina et al. 2007; Field 2007; Frayling et al. 2007; Frayling 2007; Groop 2007; Scott et al. 2007; Scuteri et al. 2007). *FTO*, also known as *FATSO*, was originally identified as one of the six genes deleted in the fused toe (*Ft*) mutant mouse (van der Hoeven et al. 1994). Heterozygous animals showed fused toes on their limbs and a thymic hyperplasia, while homozygous mice exhibited a lethal malformation of the developing brain; the embryos lost genetic control of left-right asymmetry; and, finally, the mice died around the tenth day of their embryonic development (Peters et al. 2002). The *Ft*

deletion spans several genes, of which quite a few remain of uncharacterized function. Peters and coworkers (1999) showed that one of these genes, *FTO* (*FATSO*), which is completely deleted in the *Ft* mutation, is expressed throughout embryonic development and at a high level in most organs in wild-type mice. In mouse, this novel gene spans at least 250 kb and encodes a protein of 502 amino acid residues of unknown function. It is still not known whether loss of *FTO* is a causal factor for the phenotype observed in *Ft* mutant mice. Furthermore, no deviations in BMI have been reported in *Ft* mutant mice. However, in human, unlike the associations with BMI initially reported for *GAD*, *ENPPI*, and *INSIG2*, which have not been reproduced consistently, association between the *FTO* locus and BMI is strongly supported. Frayling and coworkers (2007) studied almost 40,000 Europeans for variants of the *FTO* gene and identified an obesity risk allele. Depending on the presence of specific single nucleotide polymorphisms (SNPs) in the first intron of *FTO*, individuals weighed 1.2 to 3 kg more and had a 1.67-fold higher rate of obesity than those lacking the risk allele. Similar findings were reported by Dina et al. (2007), who studied 2,900 individuals of European ancestry, and potential Type 2 diabetes susceptibility has been correlated with another *FTO* intron 1 SNP (Scott et al. 2007).

Until recently, homologuey searches using the mouse *FTO* gene as a query only recovered sequences from vertebrates. However, with the complete genome sequencing of several marine algae, these results have been dramatically altered. While no clear homologue is found in invertebrate animals, fungi, plants, heterotrophic protists, bacteria, or archaea, we identified *FTO* homologues in the genomes of a diverse array of eukaryotic marine algae, ranging from unicellular photosynthetic picoplankton to a multicellular seaweed (Fig. 1). Specifically, *FTO* homologues were retrieved from three species within the Prasinophyceae (*Micromonas pusilla*, *Ostreococcus tauri*, and *Ostreococcus lucimarinus*) and two diatom species (*Phaeodactylum tricornutum* and *Thalassiosira pseudonana*), all of which are unicellular and which represent the only completely sequenced members of their respective lineages. Two copies of the *FTO* homologue were identified in the multicellular brown alga, *Ectocarpus siliculosus*. Furthermore, we scanned the Global Ocean Survey (GOS) dataset (Rusch et al. 2007) and recovered two additional *FTO* genes. These two sequences appear to be derived from the marine prasinophytes, due to high similarity to *FTO* homologues in the prasinophyte genomes supported by the presence of *Ostreococcus* and *Micromonas* 18S rRNA gene sequences in the same GOS sample. Strikingly, all the algae found to harbor *FTO* homologues live in marine environments, given that no *FTO* homologues were recovered from freshwater algae. We performed additional searches for *FTO* in freshwater algae using the *Chlamydomonas reinhardtii* genome sequence (Merchant et al. 2007), but to no avail. We also performed additional searches of the finished genome sequence of the red alga *Cyanidioschyzon merolae*, which thrives in acidic hot springs (Matsuzaki et al. 2004 ; Nozaki et al. 2007). Moreover, we performed these searches iteratively, using the newly discovered marine *FTO* sequences as queries, and still detected no homologues in invertebrate animals, fungi, plants, heterotrophic protists, bacteria, or archaea, confirming our initial findings.

As mentioned above, the function of *FTO* is still not known. Dina et al. (2007) detected *FTO* expression in 11 of 11 human tissue types tested, with the highest expression levels being in the hypothalamus, pituitary, and adrenal

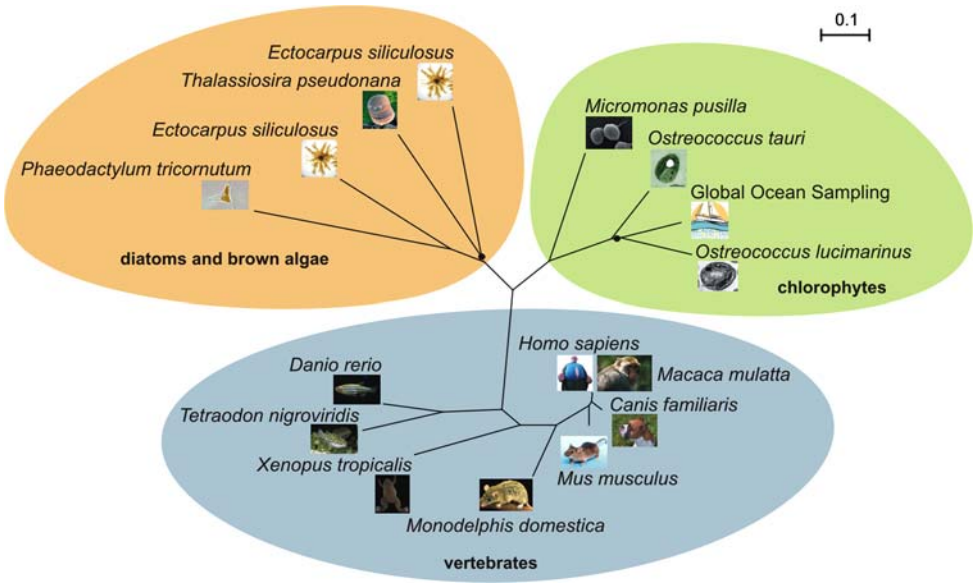

Fig. 1 Maximum likelihood tree showing the distribution of the *FTO* gene. Three major clades can be discerned: the previously described *FTO* genes in the vertebrates, the newly detected genes in diatoms and brown alga, and those of the chlorophytes and GOS sequences. All nodes are highly bootstrap supported (>70%) except two (indicated by a black circle ; 50% < BS < 70%). See Supplementary Materials for more information

glands. These findings have promoted the hypothesis that *FTO* plays a role in body weight regulation through the hypothalamic-pituitary adrenal axis. *FTO* is also expressed in rat and mouse. EST data indicate that the marine *FTO* homologues in the diatom *P. tricornutum* and the prasinophyte *M. pusilla* are expressed under standard growth conditions. Although the biological roles of the algal *FTO* homologues are still unknown, these genes can be used, together with the vertebrate sequences, to explore basic protein features. Based on primary sequence characteristics, *FTO* proteins are unlikely to be targeted to either membranes or to organelles but, rather, are predicted to be globular, cytosolic proteins with mixed α/β structures. Looking at conserved positions shows a drop from 195 positions conserved among animal sequences to only 44 conserved over all sequences, likely pinpointing the functionally essential residues. Among the most widely divergent *FTO* sequences, three amino acid residues (W, Y, and H) are strikingly overrepresented among the 44 absolutely conserved positions (see Supplementary Fig. 1). In silico predictions indicate that these residues are more likely to be located at an active site than to be at a protein-protein interface or to be surface interacting residues (Ma et al. 2003). This suggests that *FTO* may have an enzymatic function rather than be involved in protein-protein interactions. Three of the conserved positions have high prediction scores for phosphorylated residues, indicating a potential role for phosphorylation in regulation of *FTO*.

Our findings do not negate the association between *FTO* intron 1 SNPs and obesity. While identification of risk factors has advanced tremendously, for the most part, the functional ramifications of these genetic variations remain uncharacterized. In the case of *FTO*, Frayling and colleagues (2007) raised the alternative hypothesis that the intron 1 SNP might serve to alter regulation of another gene, as opposed to having a specific affect on the product encoded by *FTO*. While risk factors carry value in preventive medicine, it is mechanistic knowledge that fosters therapeutic innovation. Why marine algae harbor and express *FTO* is unclear, as is the link with obesity in humans. However, previous studies have demonstrated that algal research can be applied to investigation of vertebrate gene function. For instance, *Chlamydomonas* is often referred to as "the green yeast" because it is an easy-to-work-with eukaryotic model organism which also performs photosynthesis (see Li et al. 2004). None of the highly developed but easy to use (i.e., not involving animal work) model organisms (e.g., *Chlamydomonas*, *Arabidopsis*, yeast, *Drosophila*, and *C. elegans*) possesses an *FTO* gene. Thus, here we identify alternative systems for functional studies, such as the genetically tractable diatom *Phaeodactylum* (Siaut et al. 2007). These in turn will shed light on *FTO* function and, should that function be relevant to vertebrate homologues, thereby streamline research on genetic factors contributing to human obesity.

Methods

We initially scanned all publicly available nonredundant databases, as well as our in-house data for homologues of the mouse *FTO* gene, using BLASTP (Altschul et al. 1997). Because there was a very clear drop-off in E-value between homologues and nonhomologues (significant values, from E^{-82} to E^{-27}, then dropping to nonsignificant E-values of $\geq$0.71), selection of *FTO* homologues was straightforward. No (distantly related) genes homologous (or partially homologous) to the *FTO* genes could be identified. Next, HMMer (Eddy 1998) was used to generate a specific profile of the *FTO* gene family with hidden Markov Models, using all available sequences, and we searched NCBI EST and genome databases using TBLASTN. However, no new candidate *FTO* genes were detected.

Annotation of the *FTO* gene sequences was manually checked and corrected using ARTEMIS (Rutherford et al. 2000) when necessary. Protein sequences were aligned with CLUSTALW (Thompson et al. 1994), and after manual improvement of the alignments using BIOEDIT (Hall 1999), only 266 well-aligned positions were taken into account for tree construction. Pairwise distance trees were constructed using TREECON (Van de Peer and Wachter 1994), based on Poisson-corrected distances, while PHYML 2.4.4 (Guindon and Gascuel 2003) was used to compute the maximum likelihood tree. Bootstrap analyses with 500 replicates were performed to test the significance of the nodes. Both methods gave identical tree topologies and similar bootstrap support.

Data

Accession numbers are as follows: *Ostreococcus lucimarinus*, XP_001420808; *Ostreococcus tauri*, CAL57236; *Thalassiosira pseudonana*, jgi|Thaps3|261481|thaps1_ua_kg.chr_2000305 (http://genome.jgi-psf.org/Thaps3/Thaps3.home.html); *Phaeodactylum tricornutum*, jgi|Phatr2|41429|fgenesh1_pg.C_chr_30000044 (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html); and *Micromonas pusilla*, EU293868. FTO sequence from and *Ectocarpus siliculosus* can be obtained from the authors upon request.

Acknowledgments Sequence data of *Phaeodactylum* were produced by the Joint Genome Institute (http://www.jgi.doe.gov/). Sequence data of *Ectocarpus* were produced by Genoscope (http://www.cns.fr/). *Micromonas* culture work and genome sequencing were supported by the USDOE and a Gordon & Betty Moore Foundation grant to A.Z.W. S.R. is indebted to the Institute for

the Promotion of Innovation by Science and Technology in Flanders for a predoctoral fellowship

Note Added in Proof: After acceptance, two papers have been published on *FTO* in which this gene has been proposed to encode an Fe(II)- and 2-oxoglutarate-dependent dioxygenase (Sanchez-Pulido L, Andrade-Navarro MA (2007) BMC Biochemistry, in press, doi: 10.1186/1471-2091-8-23; Gerken et al. (2007) Science, in press, doi: 10.1126/science.1151710). The experiments of Gerken et al. (2007) showed that *FTO* is located in the nucleus, can demethylate single-stranded DNA, and its expression is regulated by feeding. However, the biologically relevant substrate of *FTO* remains unknown as well as the mechanism by which *FTO* acts on fattening.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Dina C, Meyre D, Gallina S, Durand E, Korner A, Jacobson P, Carlsson LM, Kiess W, Vatin V, Lecoeur C, Delplanque J, Vaillant E, Pattou F, Ruiz J, Weill J, Levy-Marchal C, Horber F, Potoczna N, Hercberg S, Le Stunff C, Bougneres P, Kovacs P, Marre M, Balkau B, Cauchi S, Chevre JC, Froguel P (2007) Variation in FTO contributes to childhood obesity and severe adult obesity. Nat Genet 39:724–726

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763

Field SF, Howson JM, Walker NM, Dunger DB, Todd JA (2007) Analysis of the obesity gene FTO in 14,803 type 1 diabetes cases and controls. Diabetologia 50:2218–2220

Frayling TM (2007) Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nat Rev Genet 8:657–662

Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 316:889–894

Groop L (2007) From fused toes in mice to human obesity. Nat Genet 39:706–707

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Li JB, Gerdes JM, Haycraft CJ, Fan Y, Teslovich TM, May-Simera H, Li H, Blacque OE, Li L, Leitch CC, Lewis RA, Green JS, Parfrey PS, Leroux MR, Davidson WS, Beales PL, Guay-Woodford LM, Yoder BK, Stormo GD, Katsanis N, Dutcher SK (2004) Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. Cell 117:541–552

Ma B, Elkayam T, Wolfson H, Nussinov R (2003) Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. Proc Natl Acad Sci USA 100:5772–7

Matsuzaki M, Misumi O, Shin IT, Maruyama S, Takahara M, Miyagishima SY, Mori T, Nishida K, Yagisawa F, Nishida K, Yoshida Y, Nishimura Y, Nakao S, Kobayashi T, Momoyama Y, Higashiyama T, Minoda A, Sano M, Nomoto H, Oishi K, Hayashi H, Ohta F, Nishizaka S, Haga S, Miura S, Morishita T, Kabeya Y, Terasawa K, Suzuki Y, Ishii Y, Asakawa S, Takano H, Ohta N, Kuroiwa H, Tanaka K, Shimizu N, Sugano S, Sato N, Nozaki H, Ogasawara N, Kohara Y, Kuroiwa T (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature 428:653–657

Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Maréchal-Drouard L, Marshall WF, Qu LH, Nelson DR, Sanderfoot AA, Spalding MH, Kapitonov VV, Ren Q, Ferris P, Lindquist E, Shapiro H, Lucas SM, Grimwood J, Schmutz J, Cardol P, Cerutti H, Chanfreau G, Chen CL, Cognat V, Croft MT, Dent R, Dutcher S, Fernández E, Fukuzawa H, González-Ballester D, González-Halphen D, Hallmann A, Hanikenne M, Hippler M, Inwood W, Jabbari K, Kalanon M, Kuras R, Lefebvre PA, Lemaire SD, Lobanov AV, Lohr M, Manuell A, Meier I, Mets L, Mittag M, Mittelmeier T, Moroney JV, Moseley J, Napoli C, Nedelcu AM, Niyogi K, Novoselov SV, Paulsen IT, Pazour G, Purton S, Ral JP, Riaño-Pachón DM, Riekhof W, Rymarquis L, Schroda M, Stern D, Umen J, Willows R, Wilson N, Zimmer SL, Allmer J, Balk J, Bisova K, Chen CJ, Elias M, Gendler K, Hauser C, Lamb MR, Ledford H, Long JC, Minagawa J, Page MD, Pan J, Pootakham W, Roje S, Rose A, Stahlberg E, Terauchi AM, Yang P, Ball S, Bowler C, Dieckmann CL, Gladyshev VN, Green P, Jorgensen R, Mayfield S, Mueller-Roeber B, Rajamani S, Sayre RT, Brokstein P, Dubchak I, Goodstein D, Hornick L, Huang YW, Jhaveri J, Luo Y, Martínez D, Ngau WC, Otillar B, Poliakov A, Porter A, Szajkowski L, Werner G, Zhou K, Grigoriev IV, Rokhsar DS, Grossman AR (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science 318:245–250

Nozaki H, Takano H, Misumi O, Terasawa K, Matsuzaki M, Maruyama S, Nishida K, Yagisawa F, Yoshida Y, Fujiwara T, Takio S, Tamura K, Chung SJ, Nakamura S, Kuroiwa H, Tanaka K, Sato N, Kuroiwa T (2007) A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*. BMC Biol 5:28

Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM (2006) Prevalence of overweight and obesity in the United States, 1999–2004. JAMA 295:1549–1555

Peters T, Ausmeier K, Ruther U (1999) Cloning of Fatso (Fto), a novel gene deleted by the Fused toes (Ft) mouse mutation. Mammal Genome 10:983–986

Peters T, Ausmeier K, Dildrop R, Ruther U (2002) The mouse Fused toes (Ft) mutation is the result of a 1.6–Mb deletion including the entire Iroquois B gene cluster. Mammal Genome 13:186–88

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg JF, Utterback T, Rogers YH, Falcon LI, Souza V, Bonilla-Rosso G, Eguiarte LE, Karl DM, Sathyendranath S, Platt T, Bermingham E, Gallardo V, Tamayo-Castillo G, Ferrari MR, Strausberg RL, Nealson K, Friedman R, Frazier M, Venter JC (2007) The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol 5:e77

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B (2000) Artemis: sequence visualization and annotation. Bioinformatics 16:944–945

Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-

Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316:1341–1345

Scuteri A, Sanna S, Chen WM, Uda M, Albai G, Strait J, Najjar S, Nagaraja R, Orru M, Usala G, Dei M, Lai S, Maschio A, Busonero F, Mulas A, Ehret GB, Fink AA, Weder AB, Cooper RS, Galan P, Chakravarti A, Schlessinger D, Cao A, Lakatta E, Abecasis GR (2007) Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. PLoS Genet 3:e115

Siaut M, Heijde M, Mangogna M, Montsant A, Coesel S, Allen A, Manfredonia A, Falciatore A, Bowler C (2007) Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. Gene. doi:10.1016/j.gene.2007.05.022

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Van de Peer Y, De Wachter R (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput Appl Biosci 10:569–570

van der Hoeven F, Schimmang T, Volkmann A, Mattei MG, Kyewski B, Ruther U (1994) Programmed cell death is affected in the novel mouse mutant Fused toes (Ft). Development 120:2601–2607