

# Genome-wide identification and characterization of the Hsp70 gene family in allopolyploid rapeseed (*Brassica napus* L.) compared with its diploid progenitors

## GENERAL COMMENTS

---

This work concerning Hsp70 identification and comparison in three Brassica genomes is very well conducted. Hsp70 gene family is used to understand the origin and evolution in Brassica genomes, as well as the mechanism of polyploidy adaptability. The comparison with *A. thaliana* genes is also positive to understand evolution inside diploid and tetraploid genomes. The experimental work appears to be carefully done. The authors seems to be used to this kind of exercise with other paper in the same aim but focus on another gene family (GH3, doi: 10.3390/genes10010058).

The first part of introduction is focus on the polyploidization and consequences of this process. The second part gives the example of Brassica lineage. The interest of Hsp70 is developed in the third part. The authors cited others species and it would be interested to have more information that only gene numbers, for example, in terms of divergence of functions or expressions of *A. thaliana*, rice, soybean or pepper Hsp70s. The end of the introduction define clearly the rationale and specific hypotheses of the current experiments.

The material and methods, results and discussion are well organized. Concerning results, it would have been interesting to have the percentage of similarities of genes and proteins for each sub-class (for Bn, Bo, Br and At copies) to see if there is a strong conservation and divergence between related proteins and genes. For proteins functions, Hsp70 have two important domains and only the NBD domain is really highlighted. The secondary structure analysis is also missing. Concerning gene structures with various intron-exon numbers, the possibility of alternative splicing RNA is not pointed in result (but in discussion) with may be several expressed proteins for one gene...

The main critic of this paper is the choice to name the genes. It is logical to take reference to *A. thaliana* nomenclature. But, the relationship between Bn genes and parental genomes it is not enough put forward. The authors have to change the names according to nomenclature recommended by Ostergaard and King (2008). It will be easier to see, in all the paper, homologies or differences between conserved Bn genes and parental ones, the lost or duplicated genes (Table S4 for example), in expression patterns (example: line 436, BoHsp70-5b and BnHsp70-5c homologous pair with different nomenclatures). No pseudogenes were detected?

This work has to be accepted with minor revisions.

## SPECIFIC COMMENTS

---

### **ABSTRACT:**

line 22: 47 BnHsp70 genes in *B. napus*. I think it is important to precise how many genes are on subgenome C or A.

line 25-26: sentence unclear: presence of orthologous (12 and 13) in which subfamily ?

It means also that on the 47 BnHsp70 genes only 25 are orthologous to parental genes. So, 22 are news duplicated genes in *B. napus*? If the case, that must be specified in the introduction.

That means also that 17 *B. rapa* and 7 *B. oleracea* genes were not found in *B. napus* genome and lost.

### **INTRODUCTION:**

line 52: missing space between in agriculture

At the end, the introduction needs more details :

- line 58: identification of Hsp70 genes in *B. napus* and parental genomes with a specific domain: which one ?

- line 71-84: What is the size of Hsp70 proteins? All the proteins have the two domains ?

- line 90-91:

duplicated in the corresponding genome : which ones ?

synteny between three genomes : which ones ? There are four genomes used : *B. napus* compared with *A. thaliana*, *B. oleracea* or *B. rapa* ?

- line 93: four tissues : indicate which one

### **MATERIALS AND METHODS:**

This part is described with details and is very clear to understand methods used. Some minor points to improve this part :

- In the part “identification of Hsp70 gene members” (lines 100-125), authors should identify cultivar names for reference genomes (e.g. *B. napus* Darmor bzh .....). I assume those are the cultivars that are used for tissue collection (line 182).

- line 165: the reference of the mapping method (MApInspect tool) is missing.

- In the part “plant material and tissue collection” (lines 181-186), authors should precise the developmental stage of plant with BBCH scale.

- In part “analysis of Hsp70 gene expression...”, lines 190-191 : It is possible to indicate what are juvenile stem (top of the stem ?), young leaves and the age of siliques (10- DAP as flowers ?).

- line 194: I don't understand the method to extracted the reads for each Hsp70 gene ? It was on annotated gene ontology ? If not, what were the sequences to specifically select each gene from RNA data set ?

- line 195: The method to obtain the relative gene expression levels must be more detailed.

The plant material and RNA seq approach are published by the authors in “The Gene Structure and Expression Level Changes of the GH3 Gene Family in *Brassica napus* Relative to Its Diploid Ancestors. Ruihua Wang, Mengdi Li, Xiaoming Wu, Jianbo Wang. Genes, 2019, 10(1): 58. doi: 10.3390/genes10010058. This reference must be cited and the two parts could be reduced.

- Expression in root is not followed. It is because no HSp70 have been expressed in it?

### **RESULTS:**

*genome\_wide identification of Hsp70 genes*

- line 208: “a total of 47 genes” : how many were identified from each approaches ? How many with a lack of Hsp70 specific domain? Does it mean that they are not functional? May be it could be important to indicate these genes in tables 1 and S1. No pseudogenes were identified?

- line 210: “The genes from BnHsp70-15s to BnHsp70-17s belonged to hsp110s”. Precise again that they are considered as a sub-class of Hsp70 (even if it is already mentioned in the introduction).

- Moreover, it seems to me important to compare the numbers of copies in *B. napus* genome and the parental ones. For example, no orthologous for AtHsp70-1 -3 -18 and -14 were found

in parental genomes; or still, *B. napus* genome contains only one gene for At5g02500 but 6 and 5 copies were found respectively in *B. rapa* and *B. oleracea* genomes.

- No information is given about the similarity between genes or proteins? It will be interested to know if gene or protein sequences are similar or not to At orthologous or parental copies.

- For the names given for *B. napus* genes, it will be clearer to put inside information about the membership to the sub-genome A or C. For nomenclature, as recommended by Ostergaard and King (2008), gene names were designed according to the closest At sequence homology, taking into account A or C genome location and cluster memberships. For example: BnHsp70-2 becomes BnaC.Hsp70-2. What is the *B. oleracea* copy corresponding (a to f)?

- lines 217-222: The size comparisons of Hsp70 proteins in terms of aa or kDa give the same informations. This part could be reduced.

- line 223: Authors has to mention that except BnHsp70-8s, all the proteins have negative gravity values.

- line 228: At the end of this section, authors have to compare with parental proteins as they did in lines 237-238.

#### *Phylogenetic analyses*

Why AtHsp70-1, -3, -18 and -14 were put on phylogenetic approach ? They are not detected in Bn, Bo and Br genomes?

- line 247: “subfamily E had only 5 members which were likely to be truncated”. What is the explanation? If it is the size, some others small proteins would be also designed as truncated.

- lines 253-256: “Integration analysis of the phylogenetic tree and subcellular localization prediction was similar to that described for other species, including *A.thaliana*, rice and pepper, which suggested that the Hsp70 genes have been broadly conserved in distinct taxa (Lin et al., 2001; Sarkar, Kundnani & Grover, 2013; Guo et al., 2016). ». For me, it is discussion.

- line 259: “All subfamilies except subfamily E included multiple sets of orthologs”. I think It could be nice to indicated the At orthogs for each sub-family.

- lines 262-265: Again, to clearly shown homeologs Bn gene, names have to be changed. For example, in sub-family E : BnHsp70-8b will become BnaC.Hsp7-8, BnHsp70-8a = BnaA.Hsp70-8. It would clearly show the links between Bn genes and parental ones (line 263 “12 Ar-An and 13 Co-Cn pairs”) and also if there are duplication in Bn compared to parental genomes or gene lost between parental genomes and Bn.

#### *Structure of Hsp70 genes and conserved domain of Hsp70 proteins in three Brassica species*

- line 271: and the end of the sentence, after “obtained”, put according sub-family membership.

- line 273: “The intron distribution of BnHsp70 genes was similar in the same alignment clusters”. What is the definition of a cluster ? If I look in the sub-family A the sub-group Hsp70-2, the number of introns changed. Authors have to be more precise.

- line 276: indication of the sub-category (A).

- line 298: Replace “Table S3 by Table 3, Fig.2C

- line 304: Replace “Motif 16 and 20 » by motifs 16 and 20

#### *Chromosomal distribution and duplication pattern analysis of Hsp70 genes in three Brassica species*

- line 323:” three of which formed a gene cluster” means that they are duplicated genes?

- The notion of synteny must be written in the text.

- lines 337 : reference for Schranz blocks is missing (Schranz et al., 2006).

- lines 339-341 : reference for LF, MF1 and MF2 subgenomes is missing.

- line 355: That mean “collinear regions”?

- line 377: mention of Table S5 is wrong.

*Expression patterns of Hsp70 genes in different tissues of three Brassica species*

- line 414 and line 416: “All genes except BnHsp70-6d were expressed in different tissues...”. “BnHsp70-6d lacked expression data in all samples of four tissues”. IT is not really the same idea. Authors have to change the sentences.

- line 416: “indicating most of them may play important part in the growth of these tissues”. Why the authors mentioned a role in growth tissues? It is because it was young tissues. If it is the case, it has to be mentioned.

- lines 426-428: “In addition, a similar higher expression pattern was also observed in different tissues, which was similar to Hsp70 paralogs in palaeo-polyploid soybean and pepper (Zhang et al., 2015; Guo et al., 2016).” For me, it is discussion and no result.

## **DISCUSSION**

*Genome duplications play major roles in the expansion of the BnHsp70 gene family*

- lines 497-500: concerning importance of cytosolic Hsp70s. What about the others species ? It is the same preferential retention process?

*Intron gain of BnHsp70 genes and domain loss of BnHsp70 proteins*

- line 518-519: “Also, previous researches have shown that variation of the number and placement of intron is a common process that has occurred during evolution”. Which references?

- line 526: Intron replace by intron.

## **FIGURES:**

- *Fig.1:* in the title, put *Brassica napus* in italic. Precise the cultivar (referent genome).

- *Table 1 and S1:* Precise the cultivar (referent genome) for *B. napus*, *B. rapa* and *B. oleracea*. The legends of these tables could be also supplemented briefly with the web sites used to define the characteristics.

Mw is calculated for mature proteins or with the N terminal peptide signal? Same question for pI, gravity and instability index.

- *Table S2:* Web site must be given.

-*Fig. 2:*

In the legend, for me, it is not a “phylogenetic relationship” but a dendrogram based on similarity between nucleic sequences? Why At orthologous are not indicated? Why did you put sub-families E and F together?

(A) “BnHsp70 paralogs were marked with red diamonds”. No red diamonds are present on the figure 2.

(B) The orange boxes indicate exons and black lines represent introns. TI is green boxes for exons !

Web site must be given.

On the figure, indicate in part A, the name of the sub-families (A to F). It will be clearest in comparison to Fig.1. What is the reason for the red lines (orthologs pairs)?

- *Fig S1*: I have not found a legend. N-terminal ATPase domains, EEVD, HDEL, DVIDADFTDSK, GDAWV, SPSQ (I/V) G sequences and MEME sequences motifs must be indicated. What the significance of the boxes, amino acid in green, red?

- *Table S3*: Indicate in title Bn Bo and Br Hsp70 proteins.

The authors could indicate in addition in the table the membership of the subcategories for each MEME motif and also which ones are the HSP70 specific domains.

- *Fig. 3*: The legend has to be changed. There are some results inside.

Distribution of Hsp70 gene family members on *B. rapa* (A), *B. oleracea* (B) and *B. napus* (C) chromosomes. Some genes were not shown because located on unmapped chromosomes.

The chromosome name was indicated at the top of each bar. Tandem arrays of Hsp70 genes were displayed within the blue frame. The scale of all chromosomes was in millions of base (Mb).

- *Fig.4*: The title has to be changed. Synteny is the simultaneous presence on the same chromosome of two or more loci, regardless of their genetic binding. The notion of syntenia is increasingly used to describe the preservation of the order of genes between two related species.

- Remove "Dark yellow and yellow colors represented An and Cn chromosomes of *B. napus*. Chromosomes of *B. rapa* and *B. oleracea* were represented by red and green colors, respectively. The different colors of lines indicated their syteny relationships in three species." Give the significance of trait colors inside the circus.

- The Circos is complicated to read. May be it is possible to divide it in three circos: one with only Bn genes to see homeologs genes, one with Bn sub-genome A and *B. rapa* genes, the last one with Bn sub-genome C and *B. oleracea* genes.

- *Table S4*: what is tPCK?

- *Table S6*: Abscisic acid to rewrite

- *Fig.5* : It is not necessary to put three times the color scale. The quality of fig.5B is not good compare to the two others. In the legend, authors have to give the name of the two genes missing.