

Obscuring the routes: confused data cannot reveal phylogeography of pea crop wild relatives (refutation to 'Genomic diversity and macroecology of the crop wild relatives of domesticated pea' by Smýkal *et al.* 2017)

O.E. Kosterin^{1,2*}, V.S. Bogdanova¹

¹*Institute of Cytology and Genetics SB RAS, Academician Lavrentyev ave. 10, Novosibirsk, 630090, Russia*

²*Novosibirsk State University, Pirogova str. 2, Novosibirsk, 630090, Russia*

*corresponding author; e-mail: kosterin@bionet.nsc.ru

Profound data confusion is reported for the paper by Smýkal *et al.* (Genomic diversity and macroecology of the crop wild relatives of domesticated pea. *Scientific Reports* **7**, 17384; 2017) which challenges the validity of its scientific conclusion and the data reported.

Key words: *Pisum* L., *P. sativum* L. subsp. *elatius* (Bieb.) Shmalh.; *Pisum fulvum* Sibth. et Smith; wild pea; genetic diversity; phylogeography; critics; data confusion; refutation.

Crop wild relatives can be an important source of genes useful for breeding so their genetic diversity acquires growing attention, especially in view of deterioration of their natural habitats because of human economic activity and climate change. Wild peas (*Pisum* L.) are not an exception in this respect as their genetic diversity, phylogeny and phylogeography is currently a subject of research worldwide. One of the most recent papers on the wild pea phylogeography was published in 'Scientific Reports' by Smýkal *et al.* (2017). It is based on rich materials and involves advanced modern molecular and bioinformatic methods. Unfortunately, these efforts appear to be in vain since that paper shows signs of profound data confusion so that any of its conclusions should be taken with care. This sad circumstance became possible since the confusion could be revealed only by reviewers who worked with the same accessions, although a considerable degree of sloppiness in the arrangement of supplementary data could warn a reviewer as well. We cannot judge at what level(s) the confusion took place: seeds in germplasm collection, tubes in laboratory, sequences in database, etc. Below we, however, highlight evidence that the confusion did take place. This evidence came from the data on plastid genomes of some pea accessions which we have sequenced for our forthcoming phylogenetic analysis and submitted to GenBank / European Nucleotide Archive. Eleven accessions were involved in both our study and that by Smýkal *et al.* (2017). Among them, we found only three cases of coincidence of the *trnS-trnG* plastidic spacer sequence in the genomes sequenced by us and those reported by Smýkal *et al.* (2017).

An important claim by Smýkal *et al.* (2017) was revealing introgression of plastid DNA from *Pisum sativum* L. (a diverse species including wild and cultivated forms) to *P. fulvum* Sibth. et Smith, a distinct wild species confined to the Near East. This finding referred to an intergenic spacer between *trnS* and *trnG*. The haplotypes characteristic for *P. sativum* subsp. *elatius* (Bieb.) Aschers. et Graebn. were reported to be found in six of 149 accessions of *P. fulvum*, in particular in the accession WL2140 originating from the Valley of the Cross in Jerusalem. We have got this accession in 1990s from Norman Weeden at Cornell University and recently sequenced its plastid genome (MG458702.1). However, our data did not reveal any traces of that introgression, the spacer *trnS-trnG* of WL2140, as expected, was identical to that of accession 706 (also *P. fulvum*), both obtained by us (MG458703.1) and by Smýkal *et al.* (2017) (KU678952). The *trnS-trnG* spacer haplotype reported for WL2140 by Smýkal *et al.* (2017) (KU679224) was identical to that found in the plastidic genome sequences NC_014057 and KJ806203 of the cultivated pea (*P. sativum* subsp. *sativum*), as well as to that obtained by us for J11794 (HG966675), *P. sativum* subsp. *elatius* s. l. (see below). These haplotypes differ by 5 nucleotide substitutions. Thus,

our conclusion is that at least one case of the introgression of plastid DNA of *P. sativum* into *P. fulvum* claimed by Smýkal Smýkal et al. (2017) is erroneous, and should be further confirmed in other cases.

It is not excluded that Smýkal et al. (2017) confused accessions WL2140 (*P. fulvum*) and JI2140 (*P. sativum*; traditional cultivar/landrace according to the John Innes Centre listing); the latter is mentioned in Supplementary data Table 'DARTseq-samples' worksheet but, surprisingly, not mentioned in other pages of their supplementary material (41598_2017_17623_MOESM2_ESM.xlsx).

Some other inconsistencies in the plant material assignment can be found. In page 'P_sativum-elatius' of the same supplementary table Smýkal et al. (2017) correctly stated that accession 712 (designation by Ben-Ze'ev & Zohary, 1973) is the same as L100 (designation by Herbert Lamprecht), as follows: "712 (=J13273)=L100". However, in the pages 'ITS ribotypes' and 'trnS-trnG haplotypes' they appear independently with their own NCBI numbers. Moreover, the corresponding sequences of the *trnS-trnG* spacer represent different haplotypes. This is an obvious error. We (Bogdanova et al., 2015) have sequenced the plastid genome of L100 (HG966676.1) and the *trnS-trnG* haplotype appeared identical to that of 712 (KU679018) but not of "L100" (KU679059) by Smýkal et al. (2017).

There are other discrepancies between the *trnS-trnG* spacer haplotypes reported by Smýkal et al. (2017), and the plastid genome sequences obtained by us for the same accessions. For instance, we (Bogdanova et al., 2015) have sequenced (HG966675.1) an important accession JI1794 (= 714) involved in mapping experiments (Timmerman-Vaghan et al., 1996; Weeden et al., 1998). We are confident that no confusion was involved in our study, as we revealed the inversion with breakpoints in the spacers *psal-accD* and *psbl-trnS* predicted by Palmer et al. (1985) for this pea accession. The *trnS-trnG* haplotype reported for this accession by Smýkal et al. (2017) (KU679121) differed from that revealed by us (HG966675) by 1 nucleotide substitution, but was identical to the haplotype reported by the cited authors for accessions WL2140 (KU679224) (see above) and PI344537 (KU679200).

Besides, the haplotypes of the *trnS-trnG* spacer in plastid genomes sequenced by us and reported by Smýkal et al. (2017) differed for accessions Pis2853, Pis2845, W6 26109 (provided by Petr Smýkal), JI3557 (provided by us to him) and PI344537, and coincide for accession IG64350 (provided by him to us).

The arrangement of supplementary materials (41598_2017_17623_MOESM2_ESM.xlsx) to the cited paper hints that more confusion was involved. An unexpected appearance of a cultivated pea accession JI2140 (in the paper devoted to wild peas) in one of its pages is mentioned above. All nine accessions for which "VIR, Russia" (this means N.I. Vavilov All-Russian Institute of Plant Genetic Resources, Saint-Petersburg, Russia) as indicated in the 'DARTseq-samples' page as the source have prefixes never used in VIR: 711 713, 714, 721, 722, P014, P016, P017, L100. At the same time, in the page 'ITS ribotypes', L100 was marked as obtained from JIC, while the rest and some more accessions including 712 - from 'Novosibirsk, Russia', that no doubt means from our lab. Their latter should be correct as we did exchange wild pea germplasm. This circumstance is important to rule out that the above mentioned confusions were not due to long and unknown history of accessions in germplasm collections but concerned germplasm recently and directly exchanged and at least some germplasm analysed by Smýkal *et al.* (2017) should have been derived from our stocks. Also VIR does not have pea accessions VIR10-VIR30 mentioned in the supplementary materials as *P. elatius*; most likely, the author used the ordinal numbers from the list of sent accessions (M.A. Vishnyakova, pers. comm.).

To avoid such sad cases as the paper by Smýkal et al. (2017) two things seem desirable: (i) to nominate reviewers acquainted not only in the methods but as well as the biological material involved and (ii) to thoroughly review not only the paper itself but also the arrangement of supplementary materials, which are indispensable and may the most important part of a publication as they contain the actual data.

This refutation was submitted to 'Scientific Reports' on 2.03.2017 and rejected on 12.07.2018,. In his reply Dr. P. Smýkal in particular noted that the above discrepancy could partly result from heterogeneity of accessions (not at all discussed in their paper), that VIR samples were taken from VIR Herbarium (not indicated in their paper) and did not exclude confusion of WL2140 and JI2140).

The datasets analysed during the current study are available from the GenBank / ENA repository <https://www.ncbi.nlm.nih.gov/>; the accession numbers are mentioned in the text.

Acknowledgments

The work is supported by project № 0324-2018-0018 at the Institute of Cytology & Genetics SB RAS, Novosibirsk. The authors claim absence of any conflict of interests

References

- Bogdanova VS, Zaytseva OO, Mglinets AV, Shatskaya NV, Kosterin OE, Vasiliev GV. (2015) Nuclear-cytoplasmic conflict in pea (*Pisum sativum* L.) is associated with nuclear and plastidic candidate genes encoding Acetyl-CoA carboxylase subunits. *PLoS ONE*, 10, e0119835.
- Smýkal, P., Hradilová I., Trněný, O., Brus J., Rathore A., Bariotakis M., Das R.R., Bhattacharyya D., Richards C., Coyne C.J., Pirintsos S. (2017) Genomic diversity and macroecology of the crop wild relatives of domesticated pea. *Scientific Reports*, 7, 17384.

Timmerman-Waghan, G.M., McCallum, J.A., Frew, T.J., Weeden, N. & Russel, A.C. (1996) Linkage mapping of quantitative trait loci controlling seed weight in pea (*Pisum sativum* L.). *Theor. Appl. Genet.*, 93, 431-439.

Palmer, J.D., Jorgensen, R.A. & Thompson, W.F. (1985) Chloroplast DNA variation and evolution in *Pisum*: patterns of change and phylogenetic analysis. *Genetics*, 109, 195–213.

Weeden, N., Ellis, T.H.N., Timmerman-Waghan, G.M., Swecicki, W.K., Rozov, S.M., Berdnikov, V.A. (1998) A consensus linkage map for *Pisum sativum*. *Pisum Genetics*, 30, 1-4.

Citation:

Kosterin, O.E., Bogdanova, V.S. (2018). Obscuring the routes: confused data cannot reveal phylogeography of pea crop wild relatives (refutation to ‘Genomic diversity and macroecology of the crop wild relatives of domesticated pea’ by Smýkal *et al.* 2017). *Acta Biologica Sibirica*, 4 (3), 66–68.

Submitted: 23.05.2018. **Accepted:** 05.07.2018

crossref <http://dx.doi.org/10.14258/abs.v4i3.4371>



© 2018 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).