

## Development of an InDel marker set to establish hybridization between wild boar and domestic pig (*Sus scrofa*) breeds

<sup>1,2</sup>Bendegúz Mihalik – <sup>2</sup>Krisztián Frank – <sup>2</sup>Dániel Szemethy – <sup>2</sup>Viktor Stéger – <sup>1</sup>Szilvia Kusza

<sup>1</sup> University of Debrecen, Faculty of Agricultural and Food Sciences and Environmental Management, Debrecen, Hungary

<sup>2</sup> Agricultural Biotechnology Centre, National Agricultural Research and Innovation Centre, Gödöllő, Hungary  
mihalik.bendeguz@gmail.com

### SUMMARY

Wild boar and domestic pig breeds belong to the same species (*Sus scrofa*), so they can easily have viable offspring. This could be a problem in preserving the genetic lines of wild boars, keeping clean the food industry from lower-grade hybrid boar meat, and „producing” ethically questionable trophies, too. The aim of our study was to develop a cost-efficient, fast, easy and accurate marker set which can separate the wild boars from hybrids and domestic pig breeds.

The InDel markers were developed using 59 full pig genomes of 17 different breeds (e.g. Duroc, Large White, Landrace, Mangalica, wild boar). Sequence differences between the genomes of wild boars and domestic breeds were identified in variant call files, and verified using the IGV software. Wild boar, mangalica and duroc specific primers to amplify the chosen InDel regions were designed using Primer3.

After preliminary tests five markers were chosen, three wild boar specific, one Mangalica specific and one Duroc specific one. Fluorescently labelled primers were used to make the valuation easier and more accurate with capillary electrophoresis instead of gel-electrophoresis. The markers were optimised individually and in multiplex conditions and tested in samples of 11 breeds.

In conclusion, a new, faster and cheaper set was developed to separate the wild boars from the hybrids and domestic breeds. Based on the preliminary testing on wild boars, duroc and mangalica breeds zero samples resulted false negative, so it is 100% accurate. In addition, it is a much more cost- and time-effective way than testing every single sample with STR sets.

**Keywords:** wild boar, domestic pig, *Sus scrofa*, hybridization, InDel markers

### INTRODUCTION

Domestic pig is one of the most important farm animals, providing more than hundred million tonnes of pork for meat consumption worldwide. Wild boar is the ancestor of pig, with a constantly increasing population in Hungary and across Europe (Csányi, 2014; Bieber and Ruf, 2005). Until the 1970s, domestic pigs were fed extensively by mast in forests, where they could meet with wild boars (Szabadfalvi, 1970), therefore; crossbreeding occurred easily between domestic pigs and wild boars (Babinszky et al., 2000). Nowadays, in some wild boar preserves, they were crossbred because of the domestic pig's better growth indices (reproduction rate, feed efficiency rate, etc.), degrading the “clean” wild boar populations. The wild boar-domestic pig hybrids poses a threat to the genetic diversity of boars. Also, a special phenotype with white coat colour and black patches may appear, which is noisome from a game management viewpoint and also concerning wildlife conservation. In addition, the nutrient contents of domestic pigs and hybrids also differ from wild boars (Table 1) (Vörös, 2009).

Table 1

The average nutrition values of domestic pig and wild boar meat (Vörös, 2009)

	protein %	fat %	carbohydrate %
wild boar	26	5	0.4
domestic pig	16	22.5	0.6

The annual wild boar meat production was 7177 tons in Hungary, which provides half of the whole

venison production in 2017 (Csányi et al., 2018); this volume is 1.5–2% compared to pork. Wild boar products, as other venison products, are premium quality food, thus adulteration can appear regarding such foodstuff.

DNA-based methods are one of the best ways to make clear identification of the breeds. Animals can be sampled via non-invasive methods, such as faeces and hair sampling (Taberlet et al., 1999), or from a small piece of tissue even after cooking it (Szántó et al., 2016). PCR techniques are commonly used in genetic identification, since the relatively low costs and easy use. This technique need only a small amount of tissue sample (few grams) for successful amplification (Wunderlich, 2014). Genetic markers have been developed for InDel, STR or SNP type mutations. All of these have different advantages and drawbacks (Fésüs et al., 2000). STR-s (Short Tandem Repeats) are DNA sequences made up of tandemly repeated short units not longer than six bases. STR markers are the most diverse markers, to their multi-allelic nature, thus usually fewer markers are enough for specific applications. However, this can be also their disadvantage, when the mutation rate is slightly high, which can be problematic in relationship surveys. An SNP (Single Nucleotide Polymorphism) is a mutation where one base differs in the DNA sequence of two individuals. Most of the mutations in the genome are SNPs, several millions of SNPs are present in each pig specimen (Molnár et al., 2014). Their disadvantage is that usually only two alleles are present at one locus, and their detection is more difficult; needs full sequence analyses or other DNA hybridization-based or post-amplification SNP detection method (László et

al., 2013; Molnár et al., 2014). InDel (insertion-deletion) mutations occur when longer DNA sequences are lost or gained in parts of the genome (Molnár et al., 2014). InDels are usually biallelic, the given sequence region can be either present or absent. The presence/absence of given genomic regions can be easily detected via gel- or capillary electrophoresis techniques, if the difference in size is less than or around 1000 bp (Sanders and Mason, 2016). These properties make this type of marker ideal for our research.

In the last decade, enormous efforts have been made to exploit the genome of pigs. The first pig reference genome, based on shotgun sequencing, was published in 2005 (Wernersson et al., 2005). Subsequent genome assemblies for pig and genome sequence analyses have been published continuously (Kerstens et al. 2009; Archibald et al., 2010; Zhang et al., 2010; Groenen et al., 2012; Molnár et al., 2014). These genome resources for pig, have had a huge impact on widening our knowledge about the pig genome, and are useful resources for marker development (Kerstens et al., 2009; Zsolnai et al., 2013).

In this paper, we describe InDel discovery using a large number of genomic sequences from different pig breeds and wild boars. Our aim was to develop a marker set for cost-effective, fast and easy identification

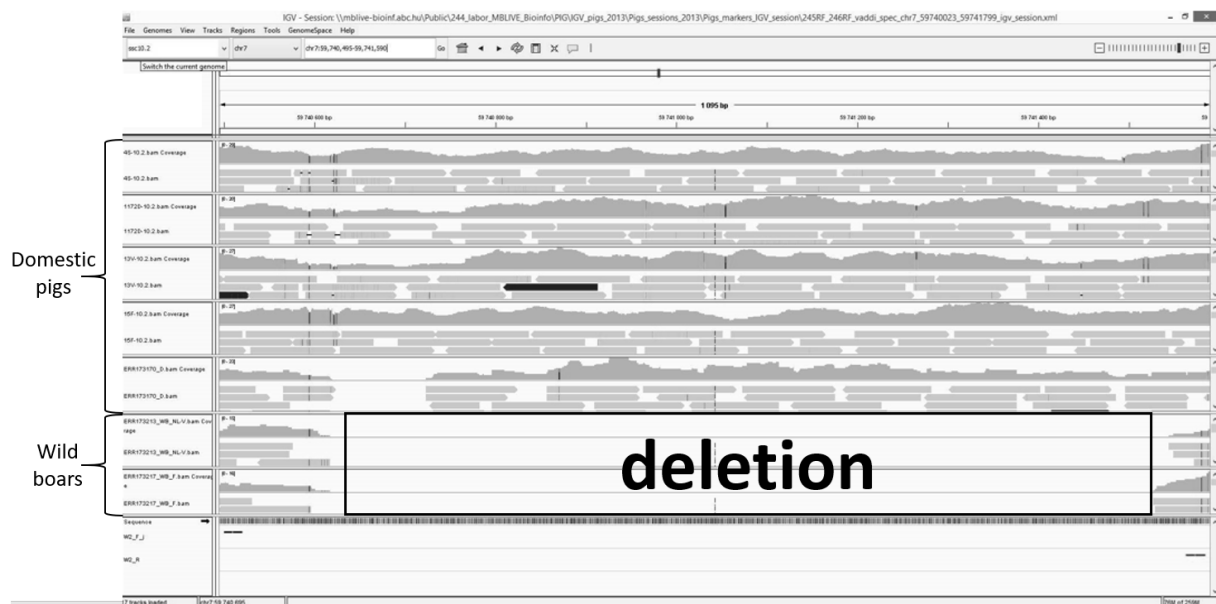
method for wild boars, domestic pigs and their hybrids. We assessed the individual and combined performance of the InDel markers, and discuss the utility of InDels as genetic markers.

**MATERIAL AND METHODS**

**Bioinformatics**

Marker development was based on published pig genome sequence data (Groenen et al., 2012; Molnár et al., 2014). In total 59 whole genomes of pig (Duroc, Hampshire, Landrace, Large White, Meishan, Pietrain, Mangalica), wild boar and other suids (bearded pig, warthog, Visayan warty pig, Javan warty pig, Sulawesi warty pig) were used in our study. Sequence differences between breeds were identified using variant call analysis (VCFtools v.0.1.16). InDel type differences were chosen, because they provide an easy and rapid way of identification with agarose gel electrophoresis. The loci showing sequence differences were checked using Integrative Genomics Viewer 2.0 software (Robinson et al., 2011), on all available breeds (*Figure 1*). The chosen InDels has to be presented in all of the actual breed’s samples but miss in other breeds. A total of 5 primers were designed with Primer3 software (<http://bioinfo.ut.ee/primer3/>) on regions showing potential diagnostic differences.

*Figure 1: IGV snapshot of the marker number 382 showing the deletion on chromosome 7 in wild boars (bottom two samples) and the absence of the deletion in other pigs (top five samples)*



**Primer selection and testing**

The designed primers were used to optimize PCR, and then pre-screen one by one on Duroc (n=4), Mangalica (n=4) and wild boar (n=4) samples. Markers showing good results on the pre-screen sample set were taken to examine other individuals. Base of the study was a sample set of 190 pigs: 120 known-breed

domestic pigs collected from breeders (12-12 duroc, pietrain, Hampshire, large white, h39 x large white, landrace, mangalica x duroc, blond mangalica, red mangalica, swallow-belly mangalica) and 70 wild boars from hunters. Total genomic DNA was isolated from hair and tissue samples using QIAamp DNA Investigator Kit (QIAGEN, Germany) and Genomic



DNA Mini Kit (Geneaid, USA) following the manufacturer's instructions. The isolated DNA samples were checked by spectrophotometry in NanoDrop ND-1000 (NanoDrop, USA) to determine the quantity and purity of the DNA. The samples were stored at -20°C until the amplification.

Final multiplex PCR amplifications were carried out in a total volume of 25 µl, containing 45 ng template DNA, primers in optimum concentration (0.20–0.40 mM) and 1 × QIAGEN Multiplex PCR Master Mix (QIAGEN GmbH, Germany), filled up to 25 µl with purified water. A Life ECO thermal cycler (Hangzhou Bioer Technology, China) was used for the amplifications with the following cycling conditions: an initial activation at 95°C for 15 min followed by 40 cycles of denaturation at 94°C for 40 s, annealing at 60°C for 40 s and extension at 72°C for 30 s with a final extension step at 72°C for 5 min. PCR products were then checked with agarose gel electrophoresis in 1,7% agarose gel, and fluorescently labelled PCR products were separated on an ABI 3100 Genetic Analyser (Applied Biosystems, USA), using LIZ 500 (Applied Biosystems, USA) as internal standard. Allele calling was performed using Peak Scanner software (Applied Biosystems, USA).

## RESULTS AND DISCUSSION

The comparative genomics analysis of 59 total pig genomes resulted in 29 genomic positions showing sequence differences between various breeds. Five out

of these genomic regions were specific for wild boar, the other were specific for different pig breeds.

The bioinformatic approach used here is a powerful technique to obtain InDel markers from genomic sequence data. Although there are microsatellite-based (Lin et al., 2013) and SNP-based analytical kits (Willkinson et al., 2012; Zsolnai et al., 2013) for distinguishing pig breeds, the InDel markers could allow a cost- and labour-efficient alternative for distinguish wild boar.

InDels were used previously for the identification of Mangalica breeds (Németh et al., 2014), as this type of markers has a number of advantageous properties, like the ability to type short amplified fragments, the relatively easy detection of variants, the low mutation rate, the potential for multiplexing markers together and a wide choice of loci to build optimum marker sets (Fondevila et al., 2012; Németh et al., 2014). This suggests the typing of InDels in simple and robust multiplexes has considerable potential in various genetic application.

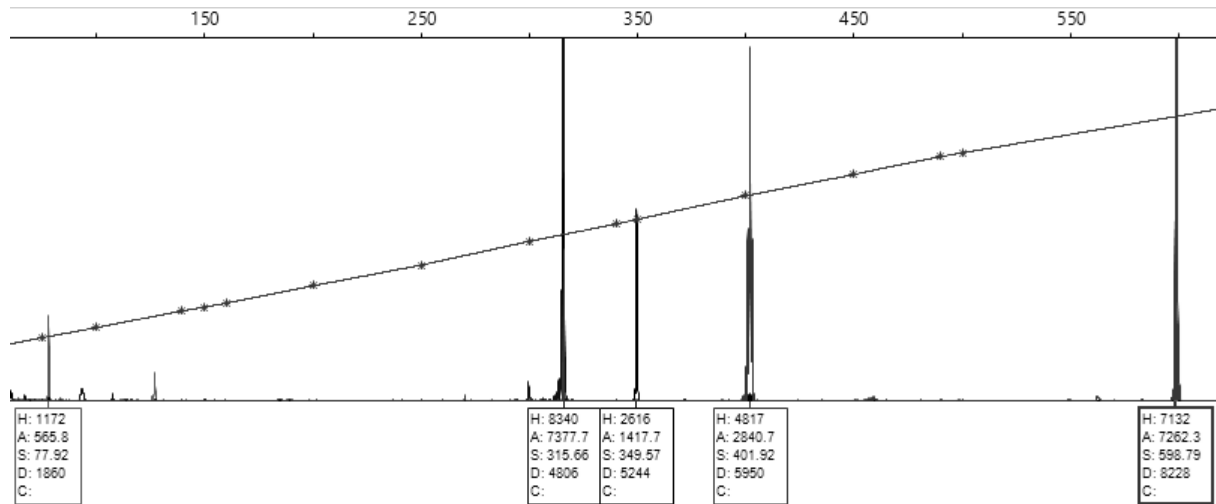
Three wild boar-specific markers (381, 382, 383) and two domestic pig-specific markers (373,379) were chosen and multiplexed to a single PCR run. In four out of these five markers, an inner primer was also used to keep product lengths under 600 bp. Fluorescently labelled primers were used for capillary electrophoresis, and in the case of inner primer the primers with two counterparts were labelled to keep colouring costs low. Primer sequences and labelling's are presented in *Table 2* and an electropherogram of an actual sample in *Figure 2*.

Table 2

The designed markers, marker types, labelling's and fragment sizes of the markers used

InDel name	Specificity	Concentration	Primers	Dye	Fragment sizes
381	Wild boar specific	10 µM	Forward Reverse	NED	350 bp / 371 bp
382	Wild boar specific	10 µM	Forward Inner reverse Reverse	PET	403 bp / 460 bp
383	Wild boar specific	10 µM	Forward Inner forward Reverse	VIC	78 bp / 86 bp
373	Duroc specific	10 µM	Forward Inner forward Reverse	PET	270 bp / 599 bp
379	Mangalica specific	10 µM	Forward Inner reverse Reverse	FAM	115 bp / 315 bp

Figure 2: The electropherogram results of a wild boar sample



## CONCLUSIONS

Genomic sequence data has been widely used to develop various markers for animal husbandry. The present study aimed to screen pig genome sequences for wild boar specific InDel markers. The developed marker set contains five InDel markers, that can be used more cost- and labour-efficient than the autosomal STR marker set described by Lin et al. (2013). Also, an InDel marker has only 2 possible fragment length, which is favourable for yes-no questions like whether a sample belongs to wild boars or not. Lack of false negative results in preliminary testing means there is no fear to identify a domestic pig as a wild boar. False positive results can be eliminated by testing all boars resulted as a hybrid with the STR set, which has still less costs than testing all samples with these markers.

These properties made it to an appropriate set to validate wild boars both in game management and food industry for genetic conservation and monitoring. After adaptation it also could be useable in controlling meat products.

In conclusion, the approach used in this study is a powerful technique to obtain InDel markers from genomic sequence data, and use the markers for species identification. Although there are microsatellite-based (Lin et al., 2013) and SNP-based analytical kits (Willkinson et al., 2012; Zsolnai et al., 2013), the InDel markers presented here could allow a cost- and labour-efficient DNA diagnostics for the identification of wild boar samples and products. This could be useful for the food industry in controlling meat products, as well as for game management purposes.

## REFERENCES

- Archibald, A. L.–Bolund, L.–Churcher, C.–Fredholm, M.–Groenen, M. A. M.–Harlizius, B.–Lee, K. T.–Milan, D.–Rogers, J.–Rothschild, M. F.–Uenishi, H.–Wang, J.–Schook, L. B.–The Swine Genome Sequencing Consortium (2010): Pig genome sequence – analysis and publication strategy. *BMC Genomics* 11: 438.
- Babinszky, L.–Csató, L.–Hancz, Cs.–Horn, P.–Horváth, L.–Kovács, J.–Szendrő, Zs.–Wittmann, M.–Zomborszky, Kovács, M. (2000): Állattenyésztés 3. – Sertés, nyúl, prémes állatok, hal. Mezőgazda kiadó, Budapest. 329 pp.
- Bieber, C.–Ruf, T. (2005): Population dynamics in wild boar *Sus scrofa*: ecology, elasticity of growth rate and implications for the management of pulsed resource consumers. *Journal of Applied Ecology* 42.6: 1203-1213.
- Csányi, S (2014) Wild boar population dynamics and management in Hungary. *Journal of Mountain Ecology* 3, 222-225.
- Csányi, S.–Márton, M.–Kovács, V.–Kovács, I.–Schally, G. (2018): Vadgazdálkodási Adattár 2017/2018-as vadászati év. Országos Vadgazdálkodási Adattár, Gödöllő. 52 pp.
- Fésűs, L.–Komlósi, I.–Varga, L.–Zsolnai, A. (2000): Molekuláris genetikai módszerek alkalmazása az állattenyésztésben. Agroinform kiadó és Nyomda Kft., Budapest. 190 pp.
- Fondevila, M.–Phillips, C.–Santos, C.–Pereira, R.–Gusmão, L.–Carracedo, A.–Butler, J. M.–Lareu, M. V.–Vallone, P. M. (2012): Forensic performance of two insertion–deletion marker assays. *International Journal of Legal Medicine* 126.5: 725-737.
- Groenen, M. A. M.–Archibald, A. L.–Uenishi, H.–Tuggle, C. L.–Takeuchi, Y.–Rothschild, M. F. et al. (2012): Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491.15: 393-398.
- Kerstens, H. H. D.–Kollers, S.–Kommadath, A.–del Rosario, M.–Dibbits, B.–Kinders, S. M.–Crooijmans, R. P.–Groenen, M. A. M. (2009): Mining for single nucleotide polymorphisms in pig genome sequence data. *BMC Genomics* 10: 4.
- László, V.–Szalai, Cs.–Pap, E.–Tóth, S.–Falus, A.–Oberfrank, F.–Szalai, Cs. (2013): Genetika és genomika. Typotex kiadó, Budapest. 188 pp.

- Lin, Y. C.–Hsieh, H. M.–Lee, J. C. I.–Hsiao, C. T.–Lin, D. Y.–Linacre, A.–Tsai, L. C. (2013): Establishing a DNA identification system for pigs (*Sus scrofa*) using a multiplex STR amplification. *Forensic Science International: Genetics* 9: 12-19.
- Molnár, J.–Nagy, T.–Stéger, V.–Tóth, G.–Marincs, F.–Barta, E. (2014): Genome sequencing and analysis of Mangalica, a fatty local pig of Hungary. *BMC Genomics* 15: 761.
- Németh, A.–Frank, K.–Bana, Á. N.–Molnár, J.–Tóth, G.–Nagy, T.–Barta, E.–Marincs, F.–Bodó, Sz.–Stéger, V. (2014): Development of Mangalica breed-specific DNA marker. 20th Youth Scientific Forum, Pannon Egyetem, Keszthely. 57-67. pp.
- Robinson, J. T.–Thorvaldsdóttir, H.–Winckler, W.–Guttman, M.–Lander, E. S.–Getz, G.–Mesirov, J. P. (2011): Integrative Genomics Viewer. *Nature Biotechnology* 29.1: 24-26.
- Sanders S. J.–Mason C. E. (2016): *Genomics, Circuits, and Pathways in Clinical Neuropsychiatry*. Academic Press, Cambridge. 796 pp.
- Szabadfalvi, J. (1970): *Az extenzív állattenyésztés Magyarországon*. Kossuth Lajos Tudományegyetem, Debrecen. 231 pp.
- Taberlet, P.–Waits, L. P.–Luikart, G. (1999): Noninvasive genetic sampling: look before you leap. *Trends in Ecology & Evolution* 14.8: 323-327.
- Vörös, G. (2009): *Amit tudnunk kell a vadhúsról*. Vadászévkönyv 2009
- Wernersson, R.–Schierup, M. H.–Jørgensen, F. G.–Gorodkin, J.–Panitz, F.–Stærfeldt, H. H.–Christensen, O. F.–Mailund, T.–Hornshøj, H.–Klein, A.–Wang, J.–Liu, B.–Hu, S.–Dong, W.–Li, W.–Wong, G. K. S.–Yu, J.–Wang, J.–Bendixen, C.–Fredholm, M.–Brunak, S.–Yang, H.–Bolund, L. (2005): Pigs in sequence space: A 0.66X coverage pig genome survey based on shotgun sequencing. *BMC Genomics* 6: 70.
- Wunderlich, L. (2014): *Molekuláris biológiai technikák*. Typotex kiadó, Budapest. 193 pp.
- Zhang, W.–Wu, W.–Lin, W.–Zhou, P.–Dai, L.–Zhang, Y.–Huang, J.–Zhang, D. (2010): Deciphering heterogeneity in pig genome assembly *Sscrofa9* by isochore and isochore-like region analyses. *PLoS One* 5.10: e13303.
- Zsolnai, A.–Tóth, G.–Molnár, J.–Stéger, V.–Marincs, F.–Jánosi, A.–Ujhelyi, G.–Koppányné Szabó, E.–Mohr, A.–Anton, I.–Szántó-Egész, R.–Egerszegi, I.–Dallmann, K.–Tóth, P.–Micsinai, A.–Brüssow, K.–Rátky, J. (2013): Looking for breed differentiating SNP loci and for a SNP set for parentage testing in Mangalica. *Archiv Tierzucht* 56.1: 200-207.

