

1 **Authentication of “mono-breed” pork products: identification of a coat colour gene marker in**  
2 **Cinta Senese pigs useful to this purpose**

3

4 L. Fontanesi <sup>a,\*</sup>, E. Scotti <sup>a</sup>, M. Gallo <sup>b</sup>, L. Nanni Costa <sup>a</sup>, S. Dall’Olio <sup>a</sup>

5

6

7 <sup>a</sup> Department of Agricultural and Food Sciences, Division of Animal Sciences, University of  
8 Bologna, Viale Fanin 46, 40127 Bologna, Italy

9 <sup>b</sup> Associazione Nazionale Allevatori Suini (ANAS), Via L. Spallanzani 4, 00161 Roma, Italy

10

11 \* Corresponding author: tel. +39 051 2096535

12 E-mail: [luca.fontanesi@unibo.it](mailto:luca.fontanesi@unibo.it) (L. Fontanesi)

13

14 **Running title:** A DNA marker useful for pork product authentication

15

16 **Highlights**

- 17 • Cinta Senese is a local Italian pig breed with a belted coat colour
- 18 • The value of its meat production chain should be protected from frauds
- 19 • We analyzed the *KIT* gene to identify a suitable DNA marker for this purpose
- 20 • A single nucleotide polymorphism (g.43597545C>T) was almost fixed in Cinta Senese
- 21 • Probability to correctly assign an unknown meat sample to Cinta Senese was ~1.00.

22 **Abstract**

23 The possibility to authenticate food products is crucial to defend local livestock production chains  
24 from frauds. Cinta Senese is an autochthonous pig breed reared under extensive or semi-extensive  
25 management systems, mainly in the Tuscany (Italy). A Protected Designation of Origin (PDO)  
26 brand for Cinta Senese meat was recently obtained. The breed is characterised by a typical black  
27 with a white-belted coat colour pattern. We analyzed a coat colour gene (*KIT*) to identify a DNA  
28 marker that could be useful for Cinta Senese meat product authentication. An informative single  
29 nucleotide polymorphism (SNP) was identified among different *KIT* gene haplotypes that were  
30 obtained from several pigs of different breeds. This SNP (g.43597545C>T; position on porcine  
31 chromosome 8 in the Sscrofa10.2 genome assembly) was genotyped by PCR-restriction fragment  
32 length polymorphism (RFLP) in 631 animals of 11 different pig breeds and one wild boar  
33 population. Allele T was almost fixed in Cinta Senese (95.9%) and absent in many breeds and was  
34 considered the tag SNP of the belted allele. Probability to correctly assign an unknown meat sample  
35 to Cinta Senese was 0.97-1.00. This DNA marker can be useful to distinguish meat of Cinta Senese  
36 pigs from meat of non-belted pigs. Thus, it could be an important tool not only to defend Cinta  
37 Senese pork chain from frauds but also to design breeding plans to eliminate non belted alleles from  
38 this pig population.

39

40 **Keywords:** SNP; coat colour gene; *KIT*; traceability; authenticity; pig breed

## 41 **1. Introduction**

42 The identification of the origin and the authentication of food products are important issues to  
43 defend livestock production chains from frauds that produce consumer distrust and undermine  
44 commercial valorisation of many local and niche products (Montowska and Pospiech, 2012).  
45 Among these products, an increasing interest during the last few years has been directed to the  
46 development of “mono-breed” labelled lines of meat and dairy products (Fontanesi, 2009). The  
47 marketing link between a breed and its products is positively considered by the consumers in terms  
48 of perceived quality and contributes to improve profitability as the products are sold at a higher  
49 price compared to undifferentiated ones. This link is mainly used to improve the economic incomes  
50 derived by local and endangered breeds that are usually less productive. The market added value is  
51 important for a sustainable exploitation of rural economies and is the fundamental driver for the  
52 conservation of endangered animal genetic resources (Fontanesi, 2009; Hoffmann, 2011).

53 Cinta Senese is an autochthonous pig breed that is reared under extensive or semi-extensive  
54 management systems, mainly in the Tuscany region (Italy). The breed is characterised by a typical  
55 black with a white belted coat colour pattern. Its origin dates back to the XIII-XIV century when  
56 belted pigs were raised in the hills around Siena as demonstrated by a famous painting of Ambrogio  
57 Lorenzetti in the Palazzo Comunale of Siena (a.D. 1340). The importance of this breed was  
58 recognized with the early constitution of the breed national herdbook that worked from the 1936 to  
59 1966 and then by a regional herdbook that was active since 1976. Just after the second world war  
60 the breed was also used to produce grey or “tramacchiati” crossbred pigs by crossing with white  
61 pigs that were fed with whey produced by the cheese factories of Pianura Padana in the North of  
62 Italy. This use was stopped by the transportation ban of the pigs due to an outbreak of diseases in  
63 1968. Since then the number of animals of this breed dropped down, almost leading to the  
64 extinction of the breed. At the end of the eighties a few projects started the recovery of this breed  
65 and in 1997 the national pig breeders herdbook preliminarily re-activated a section dedicated to  
66 Cinta Senese to promote conservation programs that made it possible to constitute a definitive

67 herdbook section for this breed in 2001 (ANAS, 2015; Franci et al., 2007). These alternate periods  
68 influenced the number of Cinta Senese heads: the number increased reaching about 160,000 in the  
69 fifties, then it decreased reaching the lowest number of 81 sows and 3 boars recorded in 1986 and  
70 after conservation programs the number of pigs raised to about 5000 heads (Franci et al., 2007;  
71 Raimondi, 1954). At present about 900 sows and 150 boars are registered in the National Herdbook  
72 (ANAS, 2015). The current stabilized number is supported by the constitution of a Protected  
73 Designation of Origin (PDO) brand for Cinta Senese meat in 2011 and the development of the Cinta  
74 Senese Consortium (Consorzio di Tutela della Cinta Senese). This consortium and the PDO  
75 contributed to the visibility of Cinta Senese products and to the added value of the meat of this  
76 breed that should be defended from potential frauds.

77 Coat colour is one of the most important traits that differentiate livestock breeds (Fontanesi,  
78 2009). DNA markers associated with coat colours in different livestock species have been already  
79 used to authenticate mono-breed dairy and meat products (D'Alessandro et al., 2007; Russo et al.,  
80 2007; Fontanesi et al., 2010, 2011).

81 As already mentioned, Cinta Senese pigs are characterised by a typical belted coat colour that  
82 can be the basis for the development of DNA markers useful for the authentication of Cinta Senese  
83 PDO products. The belted allele, in the past thought to be caused by a specific coat colour locus, is  
84 one allele of the *Dominant white* (*I*) locus series that lists several alleles derived by complex  
85 mutations in the v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (*KIT*) gene  
86 (whose product is involved in the migration of the melanoblast), namely copy number variations  
87 (CNV) and a splice site mutation (Fontanesi et al., 2010; Johansson Moller et al., 1996; Marklund et  
88 al., 1998; Pielberg et al., 2002). The combination of these mutations produces the classical white  
89 coat colour phenotype (CNV and the splice site mutation) or the patch phenotype (CNV). The  
90 molecular basis of the roan or  $I^d$  allele is not completely known even though it is due to variants  
91 affecting the *KIT* gene (Fontanesi et al., 2010). The belted allele was suggested to be derived by an  
92 uncharacterised regulatory mutation in the *KIT* gene, as it was not associated to any duplication of

93 the *KIT* gene described for other *I* alleles (Giuffra et al., 1999). Rubin et al. (2012) reported that the  
94 belted allele could be due to duplication events in the promoter region. We recently characterised  
95 different *KIT* gene haplotypes by Sanger sequencing in several cosmopolitan and local pig breeds  
96 including a few Cinta Senese pigs and identified potential breed informative haplotypes (Fontanesi  
97 et al., 2010). In this study we further analyzed the *KIT* gene and identified a DNA marker that, by  
98 comparing 11 different pig breeds and one wild boar population, was useful to design a simple  
99 genotyping test for the authentication of Cinta Senese meat.

100

## 101 **2. Materials and methods**

### 102 *2.1. Animals and DNA extraction*

103 DNA was extracted from blood samples, liver and muscle specimens and hair roots collected  
104 from a total of 602 pigs of 11 different breeds (110 Cinta Senese; 105 Italian Large White; 52  
105 Italian Landrace; 86 Italian Duroc; 32 Pietrain; 16 Hampshire; 50 Mora Romagnola; 47 Casertana;  
106 50 Apulo Calabrese; 42 Nero Siciliano; and 12 Meishan) and one wild boar population (29 animals)  
107 for a total of 631 animals. Samples were mainly obtained from previous projects (Fontanesi et al.,  
108 2010, 2014). Novel blood samples were collected during slaughtering in commercial abattoirs.  
109 DNA extraction was carried out using a standard phenol-chloroform protocol (Sambrook et al.,  
110 1989) or using the Wizard® Genomic DNA Purification kit (Promega Corporation, Madison, WI,  
111 USA), following the manufacturer instructions.

112

### 113 *2.2. PCR analysis*

114 PCR was carried out using two primer pairs. The first primer pair (forward: 5'-  
115 CCTCGCAGCAGGAGCAGT-3'; reverse: 5'-CTCAGGGCTGAGCATTCG-3') was used to  
116 amplify a fragment of 388 bp encompassing a portion of intron 17, exon 18, intron 18 and a portion  
117 of exon 19 of the porcine *KIT* gene that was used to re-sequence this gene region in 12 Cinta Senese  
118 pigs to confirm previous sequencing data obtained by Fontanesi et al. (2010). The second primer

119 pair (forward: 5'-TGAACATTGCTGACTCCCCT-3'; reverse: 5'-  
120 TGCATTTTACCTAAAGAGAAGAGC-3') was used to amplify a fragment in all 631 animals.  
121 The amplicon of 157 bp was used for the PCR-RFLP analysis described below. The amplification  
122 reactions were cycled in a 2720 Life Technologies thermal cycler (Life Technologies, Foster City,  
123 CA, USA) with the following steps: 5 min at 95 °C; 35 amplification cycles of 30 sec at 95 °C, 30  
124 sec at 56 °C, 30 sec at 72 °C; 10 min at 72 °C. The final reaction volume was of 20 µL and  
125 included: 50-100 ng of template DNA, 1 U of Taq DNA polymerase (AmpliBioTherm Taq DNA  
126 polymerase, Fisher Molecular Biology, Trevose, PA, USA; or EuroTaq DNA polymerase,  
127 EuroClone Ltd., Paington, Devon, UK); 1X PCR buffer; 2.5 mM dNTPs; 10 pmol of each primer;  
128 2.0 mM of MgCl<sub>2</sub>.

129

### 130 2.3. Sequencing and haplotype analysis

131 Amplified fragments obtained using the PCR primers of the first pair reported above were  
132 preliminarily treated with 1 µl of ExoSAP-IT<sup>®</sup> (USB Corporation, Cleveland, Ohio, USA) for 15  
133 min at 37°C. Treated amplicons were sequenced using the same PCR primers and the Big Dye v3.1  
134 cycle sequencing kit (Life Technologies, Foster City, CA, USA). Sequencing reactions were  
135 purified using EDTA 0.125 M, Ethanol 100% and Ethanol 70%, following a standard protocol.  
136 Then, the purified products were loaded on an ABI3100 Avant sequencer (Life Technologies).  
137 Obtained sequences were visually inspected and aligned with the help of the CodonCode Aligner  
138 software (<http://www.codoncode.com/aligner>) using the reference sequence of the corresponding  
139 pig *KIT* gene region (Fontanesi et al., 2010). The 28 *KIT* gene haplotypes previously reported by  
140 Fontanesi et al. (2010) from several pig breeds were aligned and compared with sequences obtained  
141 in the current study. These datasets (the previously reported *KIT* haplotypes and the additional  
142 sequences obtained here) were used to identify the most informative single nucleotide  
143 polymorphism (SNP) of the most frequent Cinta Senese haplotype.

144

145 2.4. Genotyping and data analyses

146 The genotyping protocol of the selected SNP (g.43597545C>T; position of the nucleotide  
147 coordinate on porcine chromosome 8 in the Sscrofa10.2 genome assembly of the *Sus scrofa*  
148 genome) of the Cinta Senese haplotype was based on PCR-RFLP. The amplified fragments of 157  
149 bp obtained with the second primer pair reported above was digested with the restriction enzyme  
150 *DdeI*. Briefly, restriction analysis was carried out overnight at 37 °C in a total of 13 µL of reaction  
151 volume including 5 µL of PCR product, 3 U of *DdeI* (Fermentas, Vilnius, Lithuania) and 1X  
152 reaction buffer. Produced DNA fragments were electrophoresed in 2.5% agarose gels running in  
153 TBE 1X and visualized with 1X GelRed Nucleic Acid Gel Stain (Biotium Inc., Hayward, CA,  
154 USA).

155 Hardy-Weinberg equilibrium of the genotyped SNP in the analysed populations was tested  
156 using the HWE software program (Linkage Utility Programs, Rockefeller University, New York,  
157 NY, USA). Pairwise allele and genotype frequency differences between Cinta Senese breed and all  
158 other populations were evaluated using absolute delta ( $\delta$ ) allele frequency differential. GENEPOP  
159 software version 4.0.7 (Rousset, 2008) was used to calculate population pairwise  $F_{st}$  genetic  
160 distance and G genic differentiation for each population pair (exact G test; Markov chain  
161 parameters were: Dememorisation: 10000; Batches: 100; Iterations per batch: 5000). GenAIEx 6.5  
162 software (Peakall and Smouse, 2012) was used to calculate population assignment of the pigs to the  
163 12 populations using the leave one out option. Probability to incorrectly assign an unknown meat  
164 sample to populations different from Cinta Senese or crossbred products of this breed versus all  
165 other populations (error rate) was calculated using the frequency of occurrence of Cinta Senese pigs  
166 carrying allele C. *Vice versa* probability to correctly assign an unknown meat sample to Cinta  
167 Senese ( $P_{CS}$ ) was calculated using the following formula:

168 
$$P_{CS} = 1 - (f_{TT} + f_{CT}) \quad (1)$$

169 where  $f_{TT}$  and  $f_{CT}$  are the frequency of occurrence of pigs with genotype TT or CT in the other  
170 populations.



171

### 172 **3. Results**

#### 173 *3.1. Identification and analysis of a breed-informative SNP*

174 Haplotype analyses of sequencing data reported by Fontanesi et al. (2010) who sequenced the  
175 *KIT* gene in 35 pigs of different breeds (including 6 Cinta Senese pigs) showed that putative Cinta  
176 Senese informative haplotypes could be identified (Figure 1). From these preliminary data, only two  
177 haplotypes were observed in Cinta Senese pigs (Haplotype 9 and Haplotype 24). However  
178 Haplotype 24 was identified in just one animal of this breed (in heterozygous condition with  
179 Haplotype 9). In addition, based on this information, it seemed that just a fragment of this gene,  
180 including exon 18 and part of exon 19 (Figure 2) could capture most of Haplotype 9 sequence  
181 information due to the presence of a tag marker (g.43597545C>T SNP; a synonymous SNP on exon  
182 18). Resequencing 24 other Cinta Senese haplotypes (from 12 unrelated Cinta Senese pigs), we  
183 confirmed that obtained sequences were the same as already reported for the most frequent Cinta  
184 Senese haplotype. Allele T of the tag g.43597545C>T marker was present in Cinta Senese pigs  
185 whereas allele C was present in all other haplotypes except in one rare haplotype observed in a Nero  
186 Siciliano pig (as reported from the previous data ; Fontanesi et al., 2010; Figure 1). To further  
187 validate these results and to identify a marker that could be useful to authenticate Cinta Senese meat  
188 products we set up a PCR-RFLP genotyping protocol to analyze a larger number of animals (a total  
189 of 631 pigs from different populations were genotyped). The digestion of the amplified product  
190 with *DdeI* produced two fragments of 93 + 64 bp when the amplicon contained allele T whereas  
191 when allele C was present the fragment of 157 bp remained undigested (Figure 3).

192 Allele and genotype frequencies at the g.43597545C>T SNP were obtained from 11 pig  
193 breeds including Cinta Senese and four other local Italian pig breeds (Mora Romagnola, Casertana,  
194 Apulo Calabrese and Nero Siciliano), three commercial heavy pig breeds (Italian Large White,  
195 Italian Landrace and Italian Duroc), two cosmopolitan breeds (Pietrain and Hampshire, the other  
196 belted breed included in this study), one Chinese breed (Meishan) and a European wild boar

197 population sampled in Italy (Table 1). None of the populations in which at least two alleles of the  
198 g.43597545C>T SNP were detected were in Hardy-Weinberg disequilibrium. Allele T was the most  
199 frequent in Cinta Senese pigs (95.9%). The same allele was the most frequent in the other belted  
200 breed (Hampshire) included in our survey (89.9%). In all other populations, allele T was not  
201 identified (Italian Large White, Italian Landrace, Pietrain, Mora Romagnola, Apulo-Calabrese,  
202 Meishan and European wild boars, in which only allele C was detected) or its frequency was  $\leq 6\%$   
203 (Italian Duroc, Casertana and Nero Siciliano). Comparing the frequency of allele T in Cinta Senese  
204 vs all other populations (Table 1),  $\delta$  was  $\geq 0.90$  in all comparisons ( $\delta$  was equal to 0.899 in Nero  
205 Siciliano) except against the Hampshire breed ( $\delta = 0.084$ ). The assignment test indicated that 91.8%  
206 of Cinta Senese pigs could be correctly assigned to their breed based on just the genotyped SNP  
207 whereas for all other breeds this test assigned only 25% of Hampshire pigs to the correct breed and  
208 for all other breeds assignment was 0%, mainly because the genotyped SNP was not very  
209 informative or was fixed for allele C.

210 Pairwise  $F_{st}$  measure based on the g.43597545C>T SNP indicated that all comparisons of  
211 Cinta Senese breed against all other breeds and populations were highly significant ( $P < 0.0001$ ) or in  
212 the case of Hampshire the comparison identified closeness with the Cinta Senese breed but was still  
213 significant ( $P < 0.05$ ) due to the higher frequency of allele C (0.125) in this cosmopolitan breed  
214 (Table 2). Genic differentiation for each population pair (exact G test) including Cinta Senese was  
215 highly significant for all other populations except against the Hampshire pigs confirming indirectly  
216 the results of the pairwise population matrix of mean genotypic genetic distance (Table 2).

217

### 218 *3.2. Usefulness of the KIT g.43597545C>T SNP to differentiate Cinta Senese meat*

219 Based on the genotyping data produced for the g.43597545C>T SNP it is interesting to  
220 evaluate if this *KIT* gene marker could be useful to differentiate Cinta Senese meat from meat of  
221 other pig breeds and from wild boars. In this first analysis, we will exclude the other belted breed

222 (Hampshire) investigated in this study. The Hampshire breed is not present in Italy and it should not  
223 be a problem for animals and meat coming from the same country in which Cinta Senese is raised.

224         None of the analysed Cinta Senese pigs had genotype CC. That means that if genotyping  
225 results obtained from meat of unknown origin produce genotype CC, origin from Cinta Senese can  
226 be excluded with high confidence. If we assume that there is no selection against the CC genotype  
227 in Cinta Senese pigs, the frequency of CC animals in this breed would be very low and equal to the  
228 square of the frequency of allele C, according to the formula of Hardy Weinberg equilibrium ( $0.041$   
229  $\times 0.041 = 0.00168$ ). This value would be the error rate that we might have when the genotyping test  
230 produces CC. However, we should consider that animals of this genotype, based on results obtained  
231 in this study and from our previous work (Fontanesi et al., 2010), might not have the classical  
232 belted coat colour pattern. This is because allele T can be considered the “belted allele”. Therefore,  
233 the real situation should consider a selection against the genotype CC in the Cinta Senese breed, as  
234 animals may not have the characteristic belted trait that is used to register Cinta Senese pigs as  
235 animals of this breed.

236         If we consider the case in which the genotyping result of a meat of unknown origin is CT, the  
237 error rate in assigning this meat to a hybrid pig (obtained by crossing Cinta Senese animals with  
238 genotype TT with other animals with genotype CC) different from Cinta Senese is equal to the  
239 frequency of CT pigs in the Cinta Senese population (0.089).

240         If we suppose that our comparison would only include pigs of the investigated breeds  
241 (excluding Hampshire) we should add information on the frequency of the CT genotype in all other  
242 groups of pigs (Table 1). In this case we could also consider the frequency of this genotype to  
243 define the error rate in assigning meat of unknown origin with genotype TT to a hybrid pig  
244 (obtained by crossing Cinta Senese animals with genotype TT with other animals with genotype  
245 CT) different from Cinta Senese is equal to half the frequency of the CT genotypes in the other  
246 populations, considering that crossing TT x CT, only half of the F1 animals would have genotype  
247 TT. In this way, based on a total of 505 pigs deduced from Table 1 (all genotyped animals

248 excluding Cinta Senese and Hampshire) the number of CT non-Cinta Senese pigs was 15; therefore  
249  $(15/505)/2 = 0.015$ , is the error rate for this specific question.

250 *Vice versa* probability to correctly assign an unknown meat sample to Cinta Senese ( $P_{CS}$ )  
251 obtained from the data extracted from Table 1 could be  $P_{CS} = 1 - 0.0317 = 0.9683$  that derives from  
252 the formula (1) of the Materials and methods section in which  $f_{TT} = 1/505$  and  $f_{CT} = 15/505$ , where  
253 1 is the only TT individual identified in the Duroc breed, 15 is the number of CT animals identified  
254 in Duroc (n. = 6), Casertana (n. = 4) and Nero Siciliano (n. = 5), and 505 is the total number of  
255 analysed animals, excluding Cinta Senese and Hampshire.

256 Following these procedures it is possible to easily calculate pairwise statistics for all breeds  
257 against Cinta Senese (data not shown). Considering that potential frauds would not be originated by  
258 substituting meat coming from other local breeds (that are the only ones that have the T allele,  
259 excluding the Duroc and the Hampshire breeds) but only using cheaper meat originated from  
260 commercial populations or breeds that are usually of white belted coat colour and in which allele T  
261 is not present (or it could be present at a very low frequency, that was not possible to detect in our  
262 survey), i)  $P_{CS}$  would be 1 and ii) the error rate in assigning to Cinta Senese a meat with genotype  
263 TT in case there would be doubts from its possible hybrid origin would be equal to 0.00.

264

#### 265 **4. Discussion**

266 Authentication of meat products obtained from DNA-based approaches is becoming more  
267 precise and powerful considering the large amount of genomics data that is currently available and  
268 that will be available in the future to extract useful information to answer a large number of new  
269 questions and problems arising in this field (Bertolini et al., 2015). One of the most challenging  
270 problems is the authentication of mono-breed labelled products usually obtained with local and less  
271 productive breeds (Fontanesi, 2009; Montowska and Pospiech, 2012). The problems arise by the  
272 fact that it is difficult to identify breed-specific markers as animals of different breeds can  
273 interbreed producing fertile hybrids and, for this reason, might share a large number of common

274 variants. Methods for the authentication of breed-specific products are key tools to defend the added  
275 economic value of these products that is the strategy that can obtain a sustainable conservation of  
276 local animal genetic resources, as part of an integrated development of their production chains  
277 (Fontanesi, 2009).

278 Local pig breeds represent an important resource that should be preserved and distinguished  
279 to generate economic values for niche markets that are based on their meat products. Cinta Senese  
280 products are probably one of the most valuable examples of niche pork production chain (Pugliese  
281 and Sirtori, 2012).

282 In this study, mining data obtained in genes affecting breed specific traits (in our case, coat  
283 colour), we identified a DNA marker in the *KIT* gene (g.43597545C>T) that can be useful to  
284 distinguish meat from belted pigs (the characteristic trait differentiating Cinta Senese pigs from  
285 many other pig breeds and commercial populations as well as from wild boars). Allele T was almost  
286 fixed in this breed (0.959) and for this reason can be considered its breed-characteristic allele,  
287 associated with the belted phenotype. This allele was also the most frequent in Hampshire (0.875),  
288 that is the other belted pig breed included in this study. However, at present Hampshire is not raised  
289 in Italy (only a few animals of this breed might be present in Italy), therefore there is no risk of  
290 substitution of Cinta Senese meat with meat obtained from Hampshire pigs. A potential risk could  
291 be derived by imported Hampshire meat but this is a remote possibility for the fact that pure  
292 Hampshire populations are maintained in breeding stocks for crossbreeding programs and are not  
293 commonly used for large commercial productions of final slaughtered pigs. Anyway, in this case  
294 the cost of the fraud would be high and not economically convenient.

295 Allele T was also observed in a few other breeds (Casertana, Nero Siciliano and Italian  
296 Duroc) in which it segregates at very low frequency. Only one TT animal, that apparently not had a  
297 belted phenotype, was observed in a non-belted breed. It could be possible that another very rare  
298 haplotype including allele T would be present in a few breeds or populations, as also suggested by  
299 Fontanesi et al. (2010). Moreover, the belted haplotype is also expected to segregate in non-belted

300 breeds as deduced by the results of crossbreeding programs that sometimes produce belted pigs.  
301 However, the very low frequency of allele T in other pig populations is not a big problem for the  
302 authentication of Cinta Senese products. This is due to the fact that Casertana and Nero Siciliano  
303 are local pig populations for which there is no convenience in using their meat to substitute Cinta  
304 Senese pork products and usually pure Duroc pigs are not produced on large scale, as this breed is  
305 constituted by a small nucleus used to produce sires useful for crossbreeding programs. Therefore  
306 the frequency of the CT genotype in these breeds could not be considered to calculate the error rate  
307 and  $P_{CS}$  derived by this SNP for Cinta Senese products. On the other hand, the low frequency of the  
308 alternative C allele in Cinta Senese does not substantially affect error rate and  $P_{CS}$  in this pig breed.  
309 The presence of this allele in Cinta Senese (but also in Hampshire) can indirectly confirm what is  
310 obtained from crossbreeding within the breed that sometimes produces piglets without the typical  
311 belted phenotype.

312 Other studies have investigated the possibility to authenticate mono-breed pork products or to  
313 distinguish meat of domesticated pig breeds from meat of wild boars using DNA markers. For  
314 example, markers in genes affecting coat colour (*MC1R*; *OCA2*; and *KIT*) or other phenotypic traits,  
315 like vertebral number (*NR6A1*), that are breed or population specific traits, have been already  
316 proposed to this purpose (Carrión et al., 2003; Chung and Chung, 2010; Crovetti et al., 2007;  
317 D'Alessandro et al., 2007; Fernández et al., 2004; Fontanesi et al., 2005, 2014; Okumura et al.,  
318 2000). Other approaches have used a large number of SNPs derived from the Illumina Porcine  
319 SNP60 BeadChip array (that can genotype more than 60,000 SNPs) or identified by next generation  
320 sequencing of breed specific DNA pools to identify a subset of 96 SNPs (Wilkinson et al., 2012) or  
321 193 SNPs (Ramos et al., 2011) for breed genetic discrimination among several pig breeds using  
322 different statistical approaches. Panels of microsatellites were also proposed for the same aim or to  
323 identify the level of admixture composition between different breeds (García et al., 2006; Oh et al.,  
324 2014). However the use of many markers for authentication of breed specific products has some  
325 practical and cost-limiting aspects to be solved for routine applications. A few other attempts based

326 on coat colour gene markers and multilocus microsatellite genotyping were also tested for  
327 traceability of Cinta Senese products. In particular, Crovetto et al. (2007) analysed three *MC1R*  
328 polymorphisms, one of which can distinguish Duroc pigs from black pigs (Kijas et al., 1998), and  
329 the duplication breakpoint test for the *KIT* gene that should give positive results only in white pigs  
330 (Giuffra et al., 2002). However, these markers have some limits derived by the fact that Cinta  
331 Senese breed is not fixed (or not almost fixed) for only one *MC1R* allele (Crovetto et al., 2007;  
332 Fontanesi et al., 2005). In addition, as the test for the duplication breakpoint of the *KIT* gene is  
333 designed to have an amplified product in case there is a duplication of the *KIT* gene (usually in  
334 white pigs) or absence of amplification in case there is no duplication (all other pigs with different  
335 coat colours), it could be possible that the absence of amplification might be derived by PCR failure  
336 that would prevent the correct identification of the type of tested pigs. To avoid this problem, a  
337 multiplex PCR should be designed including a control amplified fragment that should always be  
338 produced in any types of pigs (D'Alessandro et al., 2007; Fontanesi et al., 2010). Anyway, the  
339 duplication breakpoint *KIT* gene test cannot give specific indications about the breed of the pigs.  
340 Scali et al. (2012) proposed to use genotype information from 18 microsatellites to differentiate  
341 Cinta Senese pigs from Landrace, Large White, Large White x Landrace and Landrace x Cinta  
342 Senese pigs. However, their approach was not statistically supported as just 3 or 4 pigs for the other  
343 breeds or populations were included in the study that did not report the use of any standard sample  
344 to refer microsatellite allele size. In addition, more than one PCR and capillary electrophoresis  
345 might be needed to obtain the multilocus microsatellite information, even if these details were not  
346 reported in their work (Scali et al., 2012).

347 Our method based on the analysis of just one SNP (g.43597545C>T) that is highly  
348 informative for belted pigs can directly provide information from the type of pigs from which the  
349 meat is originated and is much more precise, cheaper and useful than the methods reported above  
350 based on *MC1R* and duplication breakpoint *KIT* gene analysis (Crovetto et al., 2007; Fontanesi et  
351 al., 2005) or microsatellite genotyping (Scali et al., 2012).

352

## 353 **5. Conclusions**

354 We have identified an SNP that is useful to distinguish meat of Cinta Senese pigs from meat  
355 of other non-belted pigs. This marker can be easily genotyped by PCR-RFLP using basic  
356 instruments commonly available in a molecular genetics laboratory. Therefore, it can be considered  
357 as an important tool to defend Cinta Senese pork chain from frauds. In addition, as this marker  
358 might capture the belted coat colour phenotype in pigs, it could be used to fix the belted phenotype  
359 in Cinta Senese population reducing the out-of-type animals in this breed obtained sometimes by  
360 crossing Cinta Senese pigs. Moreover, it will be interesting to evaluate if this marker could be  
361 associated with the belted phenotype in other local belted pigs that are present in Europe (i.e.  
362 Schwäbisch-Hällisches in Germany and Krškopoljski in Slovenia) and for which mono-breed  
363 products have been already proposed or could be marketed as a possible way to improve economic  
364 incomes for the farmers.

365

## 366 **Conflict of interest statement**

367 The authors declare that there are no conflicts of interest.

368

## 369 **Acknowledgements**

370 The authors would like to thank farmers and veterinarians who helped to collect samples. This  
371 project has received funding from the Italian Ministry of Agriculture, Food and Forestry (MiPAAF)  
372 under the project INNOVAGEN and from the European Union's Horizon 2020 research and  
373 innovation programme under grant agreement No 634476 (Project acronym: TREASURE). The  
374 content of this paper reflects only the author's view and the European Union Agency is not  
375 responsible for any use that may be made of the information it contains.

376

## 377 **References**



378 ANAS (2015). Registro Anagrafico. <http://www.anas.it/> (5 May 2015).

379 Bertolini, F., Ghionda, M.C., D'Alessandro, E., Geraci, C., Chiofalo, V., Fontanesi, L., 2015. A next  
380 generation semiconductor based sequencing approach for the identification of meat species in  
381 DNA mixtures. *PLoS ONE* 10, e0121701.

382 Carrión, D., Day, A., Evans, G., Mitsuhashi, T., Archibald, A., Haley, C., Andersson, L., Plastow,  
383 G., 2003. The use of MC1R and KIT genotypes for breed characterisation. *Archiv. Zootec.*  
384 52, 237-244.

385 Chung, E.-R., Chung, K.Y., 2010. Identification of Korean Native Pork using breed-specific DNA  
386 marker of KIT gene. *Kor. J. Food Sci. Anim. Res.* 30, 403-409.

387 Croveti, A., Bozzi, R., Nardi, L., Franci, O., Fontanesi, L., Russo, V., 2007. Analysis of coat colour  
388 genes for traceability of Cinta Senese products. Proceedings of 6th International Symposium  
389 on the Mediterranean Pig. October 11–13, 2007, Messina - Capo d'Orlando (ME), Italy, pp.  
390 366-368.

391 D'Alessandro, E., Fontanesi, L., Liotta, L., Davoli, R., Chiofalo, V., Russo, V., 2007. Analysis of  
392 the *MC1R* gene in the Nero Siciliano pig breed and usefulness of this locus for breed  
393 traceability. *Vet. Res. Comm.* 31 (Suppl. 1), 389-392.

394 Fernández, A., Fabuel, E., Alves, E., Rodriguez, C., Silió, L., Óvilo, C., 2004. DNA tests based on  
395 coat colour genes for authentication of the raw material of meat products from Iberian pigs. *J.*  
396 *Sci. Food Agric.* 84, 1855-1860.

397 Fontanesi, L., 2009. Genetic authentication and traceability of food products of animal origin: new  
398 developments and perspectives. *Ital. J. Anim. Sci.* 8 (Suppl. 2), 9-18.

399 Fontanesi, L., Beretti, F., Dall'Olio, S., Portolano, B., Matassino, D., Russo, V., 2011. A  
400 melanocortin 1 receptor (*MC1R*) gene polymorphism is useful for authentication of Massese  
401 sheep dairy products. *J. Dairy Res.* 78, 122-128.

402 Fontanesi, L., Bozzi, R., Tazzoli, M., Croveti, A., Davoli, R., Franci, O., Russo, V., 2005. Genetic  
403 characterization of Cinta Senese pig breed: analysis of polymorphisms in four genes affecting

404 performance and phenotypic traits. Proceedings of the International Workshop “The Role of  
405 Biotechnology for the Characterisation and Conservation of Crop, Forestry, Animal and  
406 Fishery Genetic Resources”, Villa Gualino, Turin 5-7 March 2005. Book of Proceedings, pp.  
407 175-176.

408 Fontanesi, L., D'Alessandro, E., Scotti, E., Liotta, L., Crovetto, A., Chiofalo, V., Russo, V., 2010.  
409 Genetic heterogeneity and selection signature at the *KIT* gene in pigs showing different coat  
410 colours and patterns. *Anim. Genet.* 41, 478-492.

411 Fontanesi, L., Ribani, A., Scotti, E., Utzeri, V.J., Veličković, N., Dall'Olio, S., 2014. Differentiation  
412 of meat from European wild boars and domestic pigs using polymorphisms in the *MC1R* and  
413 *NR6A1* genes. *Meat Sci.* 98, 781-784.

414 Fontanesi, L., Scotti, E., Russo, V., 2010. Analysis of SNPs in the *KIT* gene of cattle with different  
415 coat colour patterns and perspectives to use these markers for breed traceability and  
416 authentication of beef and dairy products. *Ital. J. Anim. Sci.* 9, e42.

417 Franci, O., Gandini, G., Bozzi, R., 2007. Why and how to select a local porcine breed: the case of  
418 the Cinta Senese. *Options Méditerranéennes: Série A. Séminaires Méditerranéens* 76, 13-21.

419 García, D., Martínez, A., Dunner, S., Vega-Pla, J.L., Fernández, C., Delgado, J.V., Javier Cañón, J.,  
420 2006. Estimation of the genetic admixture composition of Iberian dry-cured ham samples  
421 using DNA multilocus genotypes. *Meat Sci.* 72, 560-566.

422 Giuffra, E., Evans, G., Törnsten, A., Wales, R., Day, A., Looft, H., Plastow, G., Andersson, L.,  
423 1999. The *Belt* mutation in pigs is an allele at the *Dominant white (I/KIT)* locus. *Mamm.*  
424 *Genome* 10, 1132-1136.

425 Giuffra, E., Törnsten, A., Marklund, S., Bongcam-Rudloff, E., Chardon, P., Kijas, J.M., Anderson,  
426 S.I., Archibald, A.L., Andersson, L., 2002. A large duplication associated with dominant  
427 white color in pigs originated by homologous recombination between LINE elements flanking  
428 *KIT*. *Mamm. Genome* 13, 569-577.

429 Hoffmann, I., 2011. Livestock biodiversity and sustainability. *Liv. Sci.* 139, 69-79.

430 Johansson Moller, M., Chaudhary, R., Hellmén, E., Höyheim, B., Chowdhary, B., Andersson, L.,  
431 1996. Pigs with the dominant white coat color phenotype carry a duplication of the *KIT* gene  
432 encoding the mast/stem cell growth factor receptor. *Mamm. Genome* 7, 822-830.

433 Kijas, J.M., Wales, R., Törnsten, A., Chardon, P., Moller, M., Andersson, L., 1998. Melanocortin  
434 receptor 1 (*MC1R*) mutations and coat color in pigs. *Genetics* 150, 1177-1185.

435 Marklund, S., Kijas, J., Rodriguez-Martinez, H., Rönstrand, L., Funa, K., Moller, M., Lange, D.,  
436 Edfors-Lilja, I., Andersson, L., 1998. Molecular basis for the dominant white phenotype in the  
437 domestic pig. *Genome Res.* 8, 826-833.

438 Maudet, C., Taberlet, P., 2002. Holstein's milk detection in cheeses inferred from melanocortin  
439 receptor 1 (*MC1R*) gene polymorphism. *J. Dairy Sci.* 85, 707-715.

440 Montowska, M., Pospiech, E., 2012. Is authentication of regional and traditional food made of meat  
441 possible?. *Crit. Rev. Food Sci. Nutr.* 52, 475-487.

442 Oh, J.D., Song, K.D., Seo, J.H., Kim, D.K., Kim, S.H., Seo, K.S., Lim, H.T., Lee, J.B., Park, H.C.,  
443 Ryu, Y.C., Kang, M.S., Cho, S., Kim, E.S., Choe, H.S., Kong, H.S., Lee, H.K., 2014. Genetic  
444 traceability of black pig meats using microsatellite markers. *Asian-Austral. J. Anim. Sci.* 27,  
445 926-931.

446 Okumura, N., Kobayashi, E., Suzuki, H., Morozumi, T., Hamashima, N., Mitsuhashi, T., 2000).  
447 Breed specific mutations in *MC1R* and *KIT* genes in pigs. *Anim. Sci. J.* 8, 222-234.

448 Peakall, R., Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic  
449 software for teaching and research-an update. *Bioinformatics* 28, 2537-2539.

450 Pielberg, G., Olsson, C., Syvänen, A.C., Andersson, L., 2002. Unexpectedly high allelic diversity at  
451 the *KIT* locus causing dominant white color in the domestic pig. *Genetics* 160, 305-311.

452 Pugliese, C., Sirtori, F., 2012. Quality of meat and meat products produced from southern European  
453 pig breeds. *Meat Sci.* 90, 511-518.

454 Raimondi, R., 1954. Gli aspetti attuali della suinicoltura italiana. *L'Italia Agricola* 7, 1-16.

455 Ramos, A.M., Megens, H.J., Crooijmans, R.P., Schook, L.B., Groenen, M.A., 2011. Identification  
456 of high utility SNPs for population assignment and traceability purposes in the pig using high-  
457 throughput sequencing. *Anim. Genet.* 42, 613-620.

458 Rousset, F., 2008. Genepop'007: a complete reimplement of the Genepop software for  
459 Windows and Linux. *Mol. Ecol. Res.* 8, 103-106.

460 Rubin, C.J., Megens, H.J., Martinez Barrio, A., Maqbool, K., Sayyab, S., Schwochow, D., Wang,  
461 C., Carlborg, Ö., Jern, P., Jørgensen, C.B., Archibald, A.L., Fredholm, M., Groenen, M.A.,  
462 Andersson, L., 2012. Strong signatures of selection in the domestic pig genome. *Proc. Natl.*  
463 *Acad. Sci. USA* 109, 19529-19536.

464 Russo, V., Fontanesi, L., Scotti, E., Tazzoli, M., Dall'Olio, S., & Davoli, R., 2007. Analysis of  
465 melanocortin 1 receptor (*MC1R*) gene polymorphisms in some cattle breeds: their usefulness  
466 and application for breed traceability and authentication of Parmigiano Reggiano cheese. *Ital.*  
467 *J. Anim. Sci.* 6, 257-272.

468 Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. *Molecular cloning: a laboratory manual*. New York:  
469 Cold Spring Harbor Laboratory Press.

470 Scali, M., Vignani, R., Bigliuzzi, J., Paolucci, E., Bernini, A., Spiga, O., Niccolai, N., Cresti, M.,  
471 2012. Genetic differentiation between Cinta Senese and commercial pig breeds using  
472 microsatellite. *Electronic J. Biotechnol.* 15, 1-11.

473 Wilkinson, S., Archibald, A.L., Haley, C.S., Megens, H.J., Crooijmans, R.P., Groenen, M.A.,  
474 Wiener, P., Ogden, R., 2012. Development of a genetic tool for product regulation in the  
475 diverse British pig breed market. *BMC Genomics* 13, 580.

476

477 **Figure captions**

478

479 **Figure 1.** *KIT* gene haplotypes identified in different pig populations (D = Italian Duroc; CS =  
480 Cinta Senese; G = Gray pigs; H = Hampshire; LW = Italian Large White; M = Meishan; NS = Nero  
481 Siciliano; P = Pietrain; WB = Wild boar) as defined in Fontanesi et al. (2010) with indicated the  
482 allele of the tag marker (g.43597545C>T). Haplotypes identified in Cinta Senese pigs are  
483 evidenced.

484

485 **Figure 2.** Resequenced *KIT* gene region including the tag marker g.43597545C>T (within squared  
486 brackets) and encompassing part of intron 17, exon 18, intron 18 and part of exon 19. Exon regions  
487 are indicated in bold and evidenced. Primer regions are underlined.

488

489 **Figure 3.** PCR-RFLP patterns obtained genotyping the g.43597545C>T SNP (M = molecular DNA  
490 ladder; the genotypes are indicated above each gel line).

491

492 **Table 1.** Several statistics and data obtained for the g.43597545C>T single nucleotide  
 493 polymorphism of the *KIT* gene in different pig breeds and in wild boars.

494

Breed/population	No. of pigs	Genotypes (no. of pigs)			Allele frequencies		HWE <sup>1</sup>	$\delta^2$
		CC	CT	TT	C	T		
Cinta Senese	110	0	9	101	0.041	0.959	0.535	-
Italian Large White	105	105	0	0	1.000	0.000	-	0.959
Italian Landrace	52	52	0	0	1.000	0.000	-	0.959
Italian Duroc	86	79	6	1	0.953	0.047	0.145	0.912
Pietrain	32	32	0	0	1.000	0.000	-	0.959
Hampshire	16	0	4	12	0.125	0.875	0.449	0.084
Mora Romagnola	50	50	0	0	1.000	0.000	-	0.959
Casertana	47	43	4	0	0.957	0.043	0.673	0.916
Apulo-Calabrese	50	50	0	0	1.000	0.000	-	0.959
Nero Siciliano	42	37	5	0	0.940	0.060	0.574	0.899
Meishan	12	12	0	0	1.000	0.000	-	0.959
European wild boars	29	29	0	0	1.000	0.000	-	0.959

495

496 <sup>1</sup> Hardy Weinberg Equilibrium (P value).

497 <sup>2</sup> Absolute delta ( $\delta$ ) allele frequency differential of the T allele between Cinta Senese and all other  
 498 breed and populations.

499 **Table 2.** Pairwise population statistics comparing Cinta Senese data versus all other investigated  
500 breeds including  $F_{st}$ , genic differentiation (exact G test) and genotypic genetic distance

Breeds	$F_{st}$ (P value)	P value of the G test	Genotypic distance
Italian Large White	0.9580 (<0.0001)	<0.0001	3.755
Italian Landrace	0.9449 (<0.0001)	<0.0001	3.755
Italian Duroc	0.9089 (<0.0001)	<0.0001	3.514
Pietrain	0.9376 (<0.0001)	<0.0001	3.755
Hampshire	0.0532 (0.018)	0.2315	0.291
Mora Romagnola	0.9443 (<0.0001)	<0.0001	3.755
Casertana	0.9131 (<0.0001)	<0.0001	3.513
Apulo-Calabrese	0.9443 (<0.0001)	<0.0001	3.755
Nero Siciliano	0.9016 (<0.0001)	<0.0001	3.417
Meishan	0.9279 (<0.0001)	<0.0001	3.755
European wild boars	0.9363 (<0.0001)	<0.0001	3.755

501