Animal (2018), **12:12**, pp 2480–2488 © The Animal Consortium 2018
doi:10.1017/S1751731118000629

Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds

S. Mastrangelo[1†], M. T. Sardina[1], M. Tolone[1], R. Di Gerlando[1], A. M. Sutera[1], L. Fontanesi[2] and B. Portolano[1]

[1]*Dipartimento Scienze Agrarie, Alimentari e Forestali, Università degli Studi di Palermo, 90128 Palermo, Italy;* [2]*Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, 40127 Bologna, Italy*

(Received 25 September 2017; Accepted 23 February 2018; First published online 26 March 2018)

Runs of homozygosity (ROH) are widely used as predictors of whole-genome inbreeding levels in cattle. They identify regions that have an unfavorable effect on a phenotype when homozygous, but also identify the genes associated with traits of economic interest present in these regions. Here, the distribution of ROH islands and enriched genes within these regions in four dairy cattle breeds were investigated. Cinisara (71), Modicana (72), Reggiana (168) and Italian Holstein (96) individuals were genotyped using the 50K v2 Illumina BeadChip. The genomic regions most commonly associated with ROHs were identified by selecting the top 1% of the single nucleotide polymorphisms (SNPs) most commonly observed in the ROH of each breed. In total, 11 genomic regions were identified in Cinisara and Italian Holstein, and eight in Modicana and Reggiana, indicating an increased ROH frequency level. Generally, ROH islands differed between breeds. The most homozygous region (>45% of individuals with ROH) was found in Modicana on chromosome 6 within a quantitative trail locus affecting milk fat and protein concentrations. We identified between 126 and 347 genes within ROH islands, which are involved in multiple signaling and signal transduction pathways in a wide variety of biological processes. The gene ontology enrichment provided information on possible molecular functions, biological processes and cellular components under selection related to milk production, reproduction, immune response and resistance/susceptibility to infection and diseases. Thus, scanning the genome for ROH could be an alternative strategy to detect genomic regions and genes related to important economic traits.

Keywords: runs of homozygosity islands, genomic regions, candidate genes, local dairy cattle, bovine beadchip 50K

Implications

The genomic regions subjected to selection tend to generate runs of homozygosity (ROH) islands or hotspots. The aim of this work was to identify the differences between breeds and use the location of ROH islands to identify genes potentially involved in economically important traits. We identified several genes within ROH involved in a wide variety of biological processes, such as milk yield and composition, reproduction, immune response, resistance/susceptibility to infectious and diseases. These results showed that scanning the genome for ROH could be an alternative strategy to detect genomic regions and genes related with important economically traits.

Introduction

The development of single nucleotide polymorphism (SNP) arrays to scan the genome allow us to distinguish non-autozygotic segments that are identical by state from autozygotic and identical by descent segments (Peripolli *et al.*, 2016). A potential alternative method, called ROH, has been used in livestock for the identification of homozygous genomic regions (Purfield *et al.*, 2012; Ferenčaković *et al.*, 2013a). Runs of homozygosity are contiguous lengths of homozygous genotypes that are present in an individual because the parents transmitted identical by descending haplotypes to their offspring (Gibson *et al.*, 2006). Runs of homozygosity has been widely used as predictors of whole-genome inbreeding levels (Zhang *et al.*, 2015a; Mastrangelo *et al.*, 2016). Moreover, ROH have been used in livestock genomic studies, confirming the correlation between shared ROH and genomic regions putatively under selection (Kim *et al.*, 2013; Gaspa *et al.*, 2014; Metzger *et al.*, 2015; Szmatola *et al.*, 2016; Kukučková *et al.*, 2017; Purfield *et al.*, 2017). In fact, the genomic regions subjected to selection frequently show signatures, such as reduced nucleotide diversity, and tend to generate ROH islands or hotspots, which have high levels of homozygosity around a selected

† E-mail: salvatore.mastrangelo@unipa.it

locus compared with the rest of the genome (Szmatola *et al.*, 2016; Purfield *et al.*, 2017). Runs of homozygosity islands are not randomly distributed across the genome and are shared among individuals within a breed (Zhang *et al.*, 2015b).

A large number of cattle breeds are defined by marked phenotypic differences and, therefore, constitute valuable models to study genome evolution in response to processes such as selection and domestication. Thus, in livestock species, ROH may contribute to the detection of genomic regions that could explain phenotypic differences among breeds that affect traits of economic importance. We previously described ROH structures in three local cattle breeds (Reggiana, Cinisara and Modicana) and in Holstein cattle (Mastrangelo *et al.*, 2016). The aim of this work was to further study the distribution of ROH islands across the genome of these four cattle breeds, which may provide insights into the mechanisms underlying their genomic differences. In addition, it aimed to characterize ROH islands and identify enriched genes that could potentially explain the effects of these homozygous regions on economically important traits.

Material and methods

Samples, genotyping and data filtering

A total of 407 animals (Cinisara = 71, Modicana = 72, Reggiana = 168 and Italian Holstein = 96) were used for the analyses. All of the individuals were genotyped using the Illumina BovineSNP50 v2 BeadChip assay (Illumina Inc., San Diego, CA, USA). Single nucleotide polymorphisms were filtered to exclude loci assigned to unmapped contigs, and only those SNPs located on autosomes were considered. Quality control included call frequency $\geqslant 0.95$, minor allele frequency (MAF) $\geqslant 0.01$, and Hardy–Weinberg Equilibrium with a $P > 0.001$. SNPs that did not satisfy these quality criteria were excluded. Single nucleotide polymorphisms were mapped using the *Bos taurus* UMD 3.1.1 genome assembly.

Genetic relationship between individuals

The genetic relationship among individuals was estimated by principal components analysis (PCA) of genetic distances. This analysis was based on the identity by state (IBS) matrices of genetic distances between individuals. Principal components analysis of the genetic distance (D) matrix was performed using the multidimensional scaling option in PLINK v.1.07 (Purcell *et al.*, 2007). The graphical representation was depicted using the statistical R software (http://www.R-project.org/).

Runs of homozygosity detection

Runs of homozygosity were estimated, for each individual, using a sliding window approach of 50 SNPs in PLINK v.1.07 (Purcell *et al.*, 2007). The minimum length that constituted the ROH was set to 4 Mb. The density of the SNP panel used to generate data for ROH identification is an important factor that strongly affects autozygosity estimates. The 50K panel overestimates the number of small segments (Purfield *et al.*, 2012; Ferenčaković *et al.*, 2013b). The following criteria were used to define the ROH: (i) one missing SNP was allowed in the ROH and up to one possible heterozygous genotype, (ii) the minimum number of consecutive SNPs that constituted a ROH was set to 30, (iii) minimum density of 1 SNP every 100 kb, and (iv) maximum gap between consecutive SNPs of 1 Mb.

Identification of genomic regions and genes within runs of homozygosity

To identify the genomic regions of high homozygosity, the amount of times that each SNP appeared in the ROH was considered and normalized by dividing it by the number of animals included in the analysis. These values were plotted against the position of the SNP along the chromosome. The genomic regions were defined according to Szmatola *et al.* (2016). Adjacent SNPs having a proportion of ROH occurrences over the adopted threshold formed ROH islands. Mean linkage disequilibrium (LD) was estimated using HAPLOVIEW v. 4.2 (Barrett *et al.*, 2005) for all pairwise combinations of SNPs within each ROH island. Genomic coordinates for all identified ROH islands were also used for the annotation of genes that were fully or partially contained within each selected region using the UCSC Genome Browser (http://genome.ucsc.edu/). The genes were further analyzed with the Panther Classification System (Mi *et al.*, 2013) to identify significant ($P \leqslant 0.05$) gene ontology (GO) terms. Finally, to investigate the biological function of each annotated gene contained in ROH islands, an accurate literature search was also conducted.

Results

A PCA was used to visualize and explore the genetic relationships among breeds. The PCA (Figure 1) showed that breeds formed non-overlapping clusters and were clearly separated populations. After data quality and genetic relationships analyses, no outliers were detected.

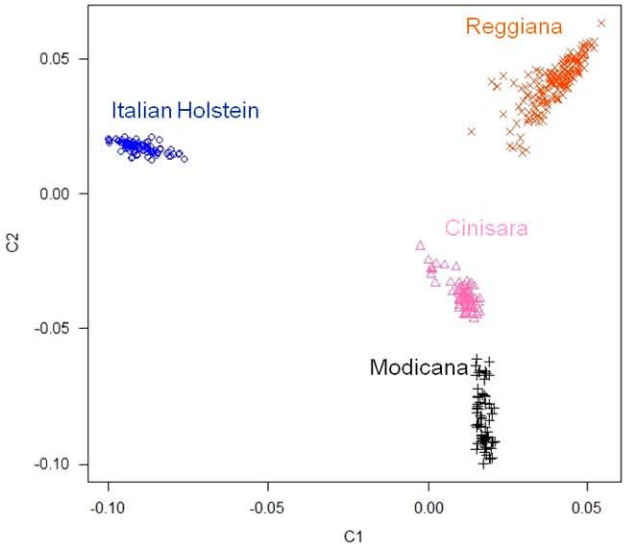

Figure 1 Genetic relationship defined with multidimensional scaling analysis for the four cattle breeds.

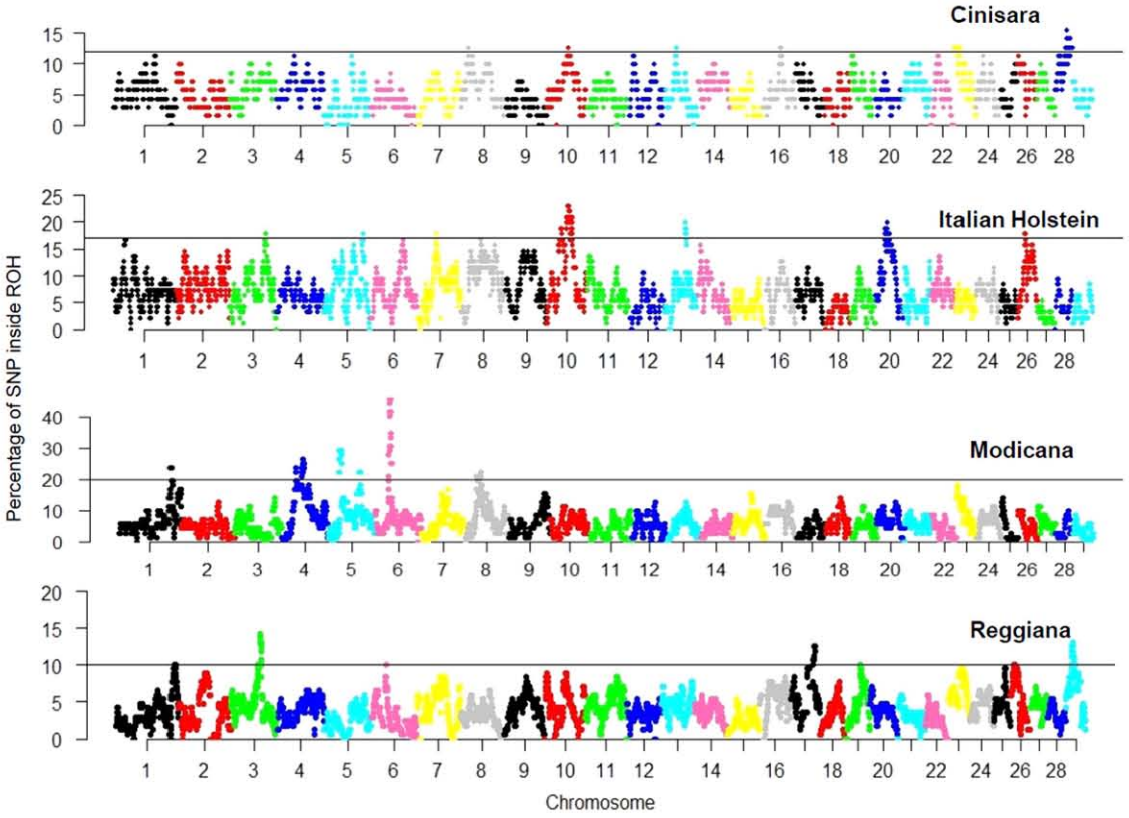

Figure 2 Genome-wide frequency of single nucleotide polymorphisms (SNPs) occurrence into runs of homozygosity (ROH) for each cattle breed. The black lines indicated the adopted threshold which defines the autozygosity islands, different per each breed (top 1% of the observations).

A total of 44 875 SNPs in Cinisara, 42 687 SNPs in Modicana, 35 270 SNPs in Reggiana, and 41 569 SNPs in Italian Holstein cattle breeds were retained after quality control for ROH detection. The top 1% of SNPs observed in the ROH was selected, and adjacent SNPs over this threshold were merged into genomic regions corresponding to ROH islands (Szmatola *et al.*, 2016). In ROH islands detected here, each SNP showed a percentage of occurrence >10% (Figure 2). This approach resulted in the identification of 11 ROH islands in Cinisara and Italian Holstein, and eight in Modicana and Reggiana (Table 1). Two overlapping ROH islands were observed between breed pairs. Modicana and Reggiana breeds showed a common genomic region on *Bos taurus* autosome (BTA) 6 (6:38 689 886 to 39 346 170 bp) and Cinisara and Italian Holstein breeds on BTA10 (10:56 464 919 to 56 792 715 bp). The genomic distribution of ROH islands was clearly non-uniform among breeds and across autosomes (Table 1). The longest ROH island was observed in Italian Holstein on BTA10 (12.42 Mb), while the shortest one was observed in Reggiana on BTA3 (0.03 Mb). BTA6 in Modicana breed had the ROH with the highest peak (Figure 2) which consisted of 38 SNPs with an occurrence in ROH >45% and a length of 2.05 Mb.

The mean r^2 value, a standard descriptive LD parameter, was estimated for all pairwise combinations of SNPs within each ROH island (Supplementary Material Table S1). In Cinisara breed, the majority of SNPs within ROH islands showed low level of LD (<0.080), and r^2 ranged from 0.024 to 0.290. The other breeds showed intermediate levels of LD within ROH islands (from 0.006 to 0.280). The highest LD level was found in the ROH island on BTA6 in Reggiana breed (0.387).

Within all of the ROH islands here reported, we identified from 126 to 347 genes (347 Italian Holstein, 250 Modicana, 190 Cinisara and 126 Reggiana). A list of genes found in the ROH islands of each breed underwent a GO enrichment analysis. Multiple categories were statistically significant ($P \leqslant 0.05$). The genes within ROH islands encompass a wide spectrum of molecular function, biological process, and cellular components. A PANTHER gene list analysis revealed a high percentage of genes involved in catalytic activity (GO:0003824), cellular processes (GO:0009987), cell part (GO:0044464), metabolic processes (GO:0008152), binding (GO:0005488) as well as biological regulations (GO:0065007) and response to stimulus (GO:0050896) in all of the ROH islands of the analyzed breeds (Table 2). Supplementary Material Table S2 provides the chromosome position, number of SNPs and number of genes per genomic region, gene symbol and full name for all of the annotated genes in each breed.

Discussion

We analyzed animals from four Italian cattle breeds with different inbreeding background and selection histories. Mastrangelo *et al.* (2016), in a previous study on evaluation on ROH in these breeds, reported the highest value of

Table 1 *List of genomic regions of extended homozygosity (ROH islands) identified in each cattle breed*

Breed	BTA	Number of SNPs	Number of genes	Start bp	End bp	Length (bp)
Cinisara	8	5	0	18 112 643	18 420 652	308 010
	10	6	1	56 464 919	56 792 715	327 797
	13	4	0	30 530 185	30 878 341	348 157
	16	23	22	43 922 935	45 552 538	1 629 604
	23	84	16	60 163	6 423 288	6 363 126
	23	7	6	10 870 036	11 251 946	381 911
	23	5	2	13 517 193	13 793 884	276 692
	28	69	49	27 655 543	32 996 400	5 340 858
	28	88	20	34 157 181	39 007 759	4 850 579
	28	9	1	39 700 262	40 191 764	491 503
	28	136	60	40 782 405	46 224 056	5 441 652
Italian Holstein	3	35	27	91 930 742	93 497 168	1 566 427
	5	2	1	99 527 745	99 569 438	41 694
	7	6	2	49 145 480	49 715 020	569 541
	10	83	71	34 907 534	40 294 545	5 387 012
	10	214	76	49 889 790	62 309 052	12 419 263
	10	47	17	63 095 461	67 118 053	4 022 593
	13	61	100	51 880 463	56 190 025	4 309 563
	20	33	11	24 266 877	26 460 587	2 193 711
	20	28	16	29 545 545	31 848 979	2 303 435
	20	27	8	34 817 221	36 570 529	1 753 309
	26	26	19	19 727 292	21 226 405	1 499 114
Modicana	1	39	25	130 168 696	132 182 348	2 013 653
	4	46	4	35 763 942	37 877 098	2 113 157
	4	113	20	51 406 099	57 744 446	6 338 348
	5	84	136	27 542 987	33 508 142	5 965 156
	5	51	18	78 776 781	82 786 530	4 009 750
	6	112	25	34 324 052	41 343 408	7 019 357
	8	53	4	29 767 566	32 749 041	2 981 476
	8	10	1	40 422 559	40 921 256	498 698
Reggiana	1	25	14	150 141 293	151 550 746	1 409 454
	1	15	3	151 736 540	152 412 536	675 997
	3	2	0	71 141 852	71 167 977	26 126
	3	71	28	73 035 441	79 378 528	6 343 088
	6	12	4	38 689 886	39 346 170	656 285
	17	97	31	56 941 968	61 788 328	4 846 361
	26	23	8	9 078 964	10 441 474	1 362 511
	29	74	28	15 819 913	23 142 122	7 322 210

inbreeding (F) based on ROH ($F_{ROH} = 0.055$) for Modicana, whereas Reggiana showed the lowest one ($F_{ROH} = 0.035$). The individuals of Italian Holstein and Reggiana showed high number of short ROH segments. Modicana and Cinisara showed similar results between them with the total length of ROH characterized by the presence of large segments due to a recent inbreeding. In this study, we reported the distribution of ROH islands across the genome of these cattle breeds to provide insights into the mechanisms underlying genomic differences among them.

Genomic regions with high frequency in runs of homozygosity

In our study, we did not perform LD pruning, but, owing to the minimum 4 Mb size of ROH segments, we tried to avoid small autozygous segments caused by LD. Indeed, a strong LD, typically extending up to ~200 kb, is common throughout the bovine genome (Mastrangelo *et al.*, 2014), and short ROH are very prevalent. To exclude these short and very common ROH, the minimum length for ROH was set to >4 Mb.

The top 1% of SNPs with the highest number of occurrence was chosen as an indication of a possible ROH island in the genome. The same threshold was reported in studies on cattle (Szmatola *et al.*, 2016) and sheep (Purfield *et al.*, 2017). Gaspa *et al.* (2014) and Sölkner *et al.* (2014) used top regions with percentage of SNP in ROH >40% within breed, whereas Mészáros *et al.* (2015) applied a threshold of 10%. Recently, a common ROH proportion higher than 7.5% was chosen as an indicator of potential autozygosity islands in

Table 2 *Gene ontology (GO) terms enriched (P < 0.05) based on runs of homozygosity islands and number of involved genes (n) for each cattle breed*

Breeds	Molecular function	Biological process	Cellular component
Cinisara	Binding (GO:0005488) $n=32$ Receptor activity (GO:0004872) $n=7$ Structural molecule activity (GO:0005198) $n=2$ Signal transducer activity (GO:0004871) $n=6$ Catalytic activity (GO:0003824) $n=40$ Transporter activity (GO:0005215) $n=9$	Cellular component organization (GO:0071840) $n=18$ Cellular process (GO:0009987) $n=57$ Localization (GO:0051179) $n=20$ Reproduction (GO:0000003) $n=5$ Biological regulation (GO:0065007) $n=36$ Response to stimulus (GO:0050896) $n=20$ Developmental process (GO:0032502) $n=15$ Immune System process (GO:0002376) $n=1$ Multicellular organismal process (GO:0032501) $n=15$ Biological adhesion (GO:0022610) $n=1$ Locomotion (GO:0040011) $n=1$ Metabolic process (GO:0008152) $n=50$ Growth (GO:0040007) $n=1$	Membrane (GO:0016020) $n=12$ Macromolecular complex (GO:0032991) n17 Cell part (GO:0044464) $n=52$ Organelle (GO:0043226) $n=32$ Extracellular region (GO:0005576) $n=12$ Synapse (GO:0045202) $n=1$
Modicana	Binding (GO:0005488) $n=37$ Receptor activity (GO:0004872) $n=20$ Structural molecule activity (GO:0005198) $n=10$ Signal transducer activity (GO:0004871) $n=17$ Catalytic activity (GO:0003824) $n=33$ Transporter activity (GO:0005215) $n=11$	Cellular component organization (GO:0071840) $n=19$ Cellular process (GO:0009987) $n=82$ Localization (GO:0051179) $n=10$ Biological regulation (GO:0065007) $n=42$ Response to stimulus (GO:0050896) $n=40$ Developmental process (GO:0032502) $n=29$ Multicellular organismal process (GO:0032501) $n=34$ Locomotion (GO:0040011) $n=11$ Biological adhesion (GO:0022610) $n=3$ Metabolic process (GO:0008152) $n=45$ Growth (GO:0040007) $n=5$ Immune system process (GO:0002376) $n=7$	Cell junction (GO:0030054) $n=2$ Membrane (GO:0016020) $n=9$ Macromolecular complex (GO:0032991) $n=29$ Extracellular matrix (GO:0031012) $n=2$ Cell part (GO:0044464) $n=57$ Organelle (GO:0043226) $n=35$ Extracellular region (GO:0005576) $n=8$
Reggiana	Binding (GO:0005488) $n=28$ Receptor activity (GO:0004872) $n=1$ Structural molecule activity (GO:0005198) $n=3$ Catalytic activity (GO:0003824) $n=30$ Transporter activity (GO:0005215) $n=6$ Signal transducer activity (GO:0004871) $n=6$	Cellular component organization (GO:0071840) $n=6$ Cellular process (GO:0009987) $n=44$ Localization (GO:0051179) $n=11$ Biological regulation (GO:0065007) $n=14$ Response to stimulus (GO:0050896) $n=11$ Developmental process (GO:0032502) $n=11$ Multicellular organismal process (GO:0032501) $n=11$ Biological adhesion (GO:0022610) $n=2$ Metabolic process (GO:0008152) $n=31$	Synapse (GO:0045202) $n=1$ Membrane (GO:0016020) $n=4$ Macromolecular complex (GO:0032991) $n=11$ Cell part (GO:0044464) $n=35$ Organelle (GO:0043226) $n=22$
Italian Holstein	Translation regulator activity (GO:0045182) $n=1$ Binding (GO:0005488) $n=74$ Receptor activity (GO:0004872) $n=12$ Structural molecule activity (GO:0005198) $n=10$ Signal transducer activity (GO:0004871) $n=8$ Catalytic activity (GO:0003824) $n=86$ Transporter activity (GO:0005215) $n=19$	Cellular component organization (GO:0071840) $n=31$ Cellular process (GO:0009987) $n=138$ Localization (GO:0051179) $n=37$ Reproduction (GO:0000003) $n=8$ Biological regulation (GO:0065007) $n=63$ Response to stimulus (GO:0050896) $n=35$ Developmental process (GO:0032502) $n=33$ Multicellular organismal process (GO:0032501) $n=20$ Biological adhesion (GO:0022610) $n=7$ Locomotion (GO:0040011) $n=7$ Metabolic process (GO:0008152) $n=112$ Immune system process (GO:0002376) $n=10$	Synapse (GO:0045202) $n=2$ Cell junction (GO:0030054) $n=3$ Membrane (GO:0016020) $n=17$ Macromolecular complex (GO:0032991) $n=39$ Cell part (GO:0044464) $n=121$ Organelle (GO:0043226) $n=64$ Extracellular region (GO:0005576) $n=8$

cattle (Kukučková *et al.*, 2017). Therefore, we have employed a stricter criteria compared with the last two works mentioned above.

The ROH peaks were distributed and shared among individuals, and it was clear that they were signs of common ROH islands within breeds. Some of these genomic regions overlapped with ROH islands found in other studies (Table 3). The ROH islands reported on BTA4 and BTA5 in Modicana overlapped with ROH islands reported in Pinzgau (Kukučková *et al.*, 2017) and Simmental (Szamatola *et al.*, 2016). Several studies (Sölkner *et al.*, 2014; Mészáros *et al.*, 2015; Szamatola *et al.*, 2016; Kukučková *et al.*, 2017) showed ROH islands located on BTA6. These regions overlapped with the ones obtained in our study for Modicana (34.32 to 41.34 Mb) and Reggiana (38.69 to 39.35 Mb). Sölkner *et al.* (2014) studying Taurine and Indicine cattle breeds, identified a region in BTA16 (43.80 to 44.97 Mb) visible only in Taurine. This overlapped with a region obtained in our study in

Table 3 *Comparison among overlapped runs of homozygosity (ROH) islands here detected and those reported in previous studies*

Breed	BTA	Position (Mb)	References
Pinzgau	4	52.42 to 65.05	Kukučková *et al.* (2017)
Modicana	4	51.41 to 57.74	This study
Simmental	5	78.71 to 80.94	Szmatola *et al.* (2016)
Modicana	5	78.78 to 82.79	This study
Pinzgau	6	35.46 to 42.31	Kukučková *et al.* (2017)
Tyrol Grey	6	36.28 to 41.12	Mészáros *et al.* (2015)
Modicana	6	34.32 to 41.34	This study
Simmental	6	38.34 to 40.10	Szmatola *et al.* (2016)
Taurine	6	38.27 to 39.45	Sölkner *et al.* (2014)
Reggiana	6	38.69 to 39.35	This study
Taurine	16	43.80 to 44.97	Sölkner *et al.* (2014)
Red Polish	16	43.52 to 46.19	Szmatola *et al.* (2016)
Simmental	16	42.89 to 46.77	Szmatola *et al.* (2016)
Limousin	16	43.37 to 46.07	Szmatola *et al.* (2016)
Cinisara	16	43.92 to 45.55	This study
Holstein	20	28.33 to 32.29	Szmatola *et al.* (2016)
Italian Holstein	20	29.54 to 31.85	This study
Holstein	20	34.47 to 35.48	Szmatola *et al.* (2016)
Italian Holstein	20	34.82 to 36.57	This study
Italian Holstein	26	21.15 to 23.00	Gaspa *et al.* (2014)
Italian Holstein	26	19.73 to 21.23	This study
Simmental	28	39.77 to 40.57	Szmatola *et al.* (2016)
Cinisara	28	39.70 to 40.19	This study

Cinisara (43.92 to 45.55 Mb). Similar results were also reported for Red Polish, Simmental, and Limousine cattle breeds (Szmatola *et al.*, 2016). Runs of homozygosity islands identified on BTA20 in Italian Holstein were also described by Szamatola *et al.* (2016) in Holstein. Moreover, the peak identified on BTA26 in Italian Holstein partially overlapped with those obtained for the same breed by Gaspa *et al.* (2014) in which the stearoyl-CoA desaturase (*SCD*) locus is located. Finally, the ROH islands reported on BTA28 in Cinisara overlapped with an ROH island reported in Simmental (Szamatola *et al.*, 2016). These results suggested that some of the ROH islands are common among different cattle breeds, and harbor variants that are undergoing selection independently of production and selection characteristics (Szmatola *et al.*, 2016). The inconsistencies among the criteria defining ROH islands makes it difficult to compare studies because the lack of consensus allows different thresholds and thus different signals (Peripolli *et al.*, 2016). However, the overlapping ROH islands among studies provided good evidence that they are not artifacts but genuine genomic regions affected by inbreeding. It is important to highlight that ROH islands can also be partly explained by the reduced recombination rate. Indeed, despite ROH being more or less equally distributed throughout the chromosomes, ROH islands were mostly found in regions with low recombination rates (Purfield *et al.*, 2017). To verify this distribution in our cattle breeds and to determine if recombination rate impacted ROH islands, the linkage information published by the USDA (Ma *et al.*, 2015) were considered. Ma *et al.* (2015) assessed the relationship between recombination rate and chromosomal locations because recombination rates are known to differ considerably across chromosomal locations, including telomeres and centromeres. All cattle autosomes are acrocentric with the centromere located at the beginning and the telomere at the end of the chromosome. These authors reported a very low recombination rate near the centromere and the beginning of each chromosome, and they showed that the middle of the chromosome had a decreased recombination rate, although the centromere is far. Moreover, Ma *et al.* (2015) highlighted that this low recombination rate in the middle of chromosome was not universal across all bovine chromosomes, but more pronounced for some of them (i.e. BTA9, 10, 11, 13, 15, 16, 19 and 23). Following the smooth spline plotting of the recombination rate reported by Ma *et al.* (2015) in cattle, we checked if the ROH islands shown in Table 1 overlapped with regions of the genome showing low recombination rates. Some ROH islands actually overlapped with regions having low recombination rates (Table 1) as reported in previous studies in sheep (Purfield *et al.*, 2017). Moreover, a previous study on cattle (Purfield *et al.*, 2012) reported a correlation between extensive LD and high incidence of ROH. The majority of SNPs within ROH islands showed similar LD levels as those computed for the entire chromosome (Supplementary Material Table S2), with the exception of two ROH islands (on BTA16 in Cinisara and on BTA 6 in Reggiana). Therefore, their existence was not easy explained on the basis of just LD (Nothnagel *et al.*, 2009).

Identification of candidate genes within runs of homozygosity

We found that some SNPs occurred in regions of poor gene content. Some of the identified ROH islands, such as on BTA10 in the Cinisara breed, contained only one annotated gene (*WDR72*) or uncharacterized genes (i.e. *LOC107132862*). This may reflect selection acting on uncharacterized regulatory regions or simply the fixation of non-coding DNA by genetic drift due to the absence of any selection (Qanbari *et al.*, 2011). An enrichment of genes involved in several GO-terms was observed in the four cattle breeds. We have not discussed in detail all of the genomic regions associated with ROH islands. Instead, we focused on selected genes in highly GO-enriched terms that, on the basis of the literature, showed associations with several specific traits related to livestock. Therefore, the functions of candidate genes within ROH islands play important roles in cattle and other livestock species are summarized for each breed.

In Cinisara, the ROH islands were identified on BTA8, 10, 13, 16, 23 and 28. A total of 40 genes were identified as being related to catalytic activity (GO:0003824), with genes implicated in immune response and immune regulation (*PIK3CD* and *SPSB1*, respectively) (Ramey *et al.*, 2013). A high number of genes ($n = 57$) were identified as being related to cellular process (GO:0009987). Among these, some candidate genes mapped on BTA16, such as *PEX14*, which is related to dairy production, *KIF1B*, which is under

strong selection in dairy Holstein cattle (Flori *et al.*, 2009), and *RERE*, which is implicated in embryonic growth and reproductive development (Ramey *et al.*, 2013). Moreover, 52 identified genes were also related to cell part (GO:0044464) in which we highlighted the *ADK* gene on BTA28, which is involved in a physiological state (Ramey *et al.*, 2013). Other candidate genes within the ROH islands on BTA28 were *NRG3* and *PPYR1*, which are related with bovine mammary gland development and milk production, respectively (Ogorevc *et al.*, 2009).

The ROH islands in the Modicana breed were identified on BTA1, 4, 5, 6, and 8. A total of 37, 82 and 69 genes were identified as being related to binding (GO:0005488), cellular process (GO:0009987) and cell part (GO:0044464), respectively. Several enriched GO-terms contained genes related with milk production, such as the *LALBA* gene, a major whey protein that showed a significant association with the milk protein profile (Huang *et al.*, 2012). On BTA6, the most homozygous region (>45% of individuals having the ROH island) was found (6:37 019 972 to 39 069 719 bp) and it contained an intriguing element. A quantitative trait locus (QTL) on this chromosome affecting milk fat and protein concentrations has been reported (Zhang *et al.*, 1998). The QTL, containing six genes (*ABCG2*, *PKD2*, *SPP1*, *MEPE*, *IBSP* and *LAP3*), was identified within one ROH island in our study. In this chromosomal region, several genes associated with milk production traits are annotated, such as *FAM13A1*, a gene near a milk protein QTL related to the protein content (Cohen *et al.*, 2014). The *ABCG2* gene harbors a quantitative trait nucleotide for milk composition in cattle (Olsen *et al.*, 2008). *LAP3* has been associated with milk production traits (Zheng *et al.*, 2011) and with calving ease in dairy cows (Olsen *et al.*, 2008). Other interesting candidate genes, based on their suggested molecular function, were found, such as *CAV1* and *CAV2* on BTA4, which are implicated in the immune system (Qanbari *et al.*, 2014). On BTA5, the *KRT* gene family, which is associated with epithelial development, was highlighted, together with *TFCP2*, which contained a QTL associated with fertility (Moore *et al.*, 2016). Moreover, we observed a genomic region within the ROH island that contained olfactory receptor family genes. Olfactory receptors detect and identify a wide range of odors and chemosensory stimuli, a necessity for finding food, detecting mates and offspring, recognizing territories and avoiding danger. They are also reported to be duplicated within the bovine genome, suggesting that they may be under strong selection for newly evolving functions (Qanbari *et al.*, 2014).

In the Reggiana breed, the ROH islands were identified on BTA1, 3, 6, 17, 26 and BTA29. A total of 30 genes were identified as being related to catalytic activity (GO:0003824) with candidate genes, such as *DIRAS3*, which is involved in reproductive traits (Cheng *et al.*, 2007), and *PTEN*, which is involved in mammary gland function (Li *et al.*, 2015). A total of 44 genes were identified as being related to cellular process (GO:0009987) with genes, such as *MINPP1*, which is associated with milk fatty acid traits in dairy cattle (Li *et al.*, 2015). Moreover, 35 genes identified in ROH islands were related to cell part (GO:0044464), and several of these genes (such as *TAOK3* and *NCAPG*) have been previously associated with milk production traits (Li *et al.*, 2010; Weikard *et al.*, 2012). We also detected the *SLC35D1* gene, which is associated with the immune system (Qanbari *et al.*, 2014), and the *KSR2* gene, which affects milk production traits (Pimentel *et al.*, 2011).

In the Italian Holstein breed, the 11 ROH islands were found on BTA3, 5, 7, 10, 13, 20 and BTA26. The most representative supercluster identified by the GO-term enrichment contained several candidate genes involved in milk production traits and reproduction. Among the 86 genes related to catalytic activity (GO:0003824), we found *RHOV*, which is related to dairy production (Gutiérrez-Gil *et al.*, 2015), on BTA10 and *SCD*, which has a large influence on milk fat composition because it plays a major role in determining the monounsaturated fatty acids, primarily oleic acid and the CLA content of milk fat (Rincon *et al.*, 2012), on BTA26. The greatest number of genes (138) was related to cellular process (GO:0009987). Of these, *BMP4* is involved in the development and functioning of follicles (Qanbari *et al.*, 2010), and *OXT* and *AVP* play major roles in regulating estrous behavior in dairy cows (Kommadath *et al.*, 2011). Finally, 121 genes related to cell part (GO:0044464) were also found. The *PELO* gene on BTA20 is involved in dairy production (Gutiérrez-Gil *et al.*, 2015). Other important candidate genes within the ROH island were *DIO1*, which is related to milk synthesis and energy metabolism (Connor *et al.*, 2003) on BTA3 and *SLC2A4RG*, which is involved in lactation persistency (Nayeri *et al.*, 2016), on BTA10. *C9*, which is involved in immune response, was located within a QTL region for mastitis-related traits (Sahana *et al.*, 2013).

As reported above, several enriched GO-terms were related to milk production, reproduction, immune response, and resistance/susceptibility to infections and diseases. This indicated that the analyzed individuals may have experienced selective pressure on their genomes for these specific traits. Some genomic regions may be fixed in individuals within a population as a result of artificial or natural selection for reasons such as adaptability or productivity. Cinisara and Modicana are two breeds that have excellent abilities to adapt to harsh environments, high resistance levels to infections and diseases, good maternal aptitudes, and high-quality milk production. Genes that are involved in these traits were detected in our study using the ROH approach and were consistent with the phenotypic characteristics of these two breeds. Recently, a study on local sheep breeds (Mastrangelo *et al.*, 2017) revealed the presence of ROH islands in genomic regions that harbor candidate genes for selection in response to environmental stress and which underlie local adaptation. The presence of many immune system-related genes in the identified ROH islands could reflect selection (natural or artificial) for disease resistance. Reggiana and Italian Holstein are two breeds reared and selected for milk production, and in accordance with this phenotypic trait, our results emphasized the presence of dairy-related genes within the ROH islands. Currently, in

dairy cattle, such as Holstein, the systemic decline of fertility is being observed, in agreement with the several genes implicated in affecting the reproductive traits highlighted in this work. Kim *et al.* (2013) found that several genomic regions within ROH were associated with economically important traits, including milk, fat and protein yields. Therefore, the annotated genes that mapped to these ROH islands were perceived as exposed to selection.

Conclusion

In this work, we examined the distributions of ROH islands across the genomes of four cattle breeds with similar production aptitudes but different selection histories. We confirmed that the ROH islands were clearly non-uniform among breeds and across chromosomes. In fact, different ROH islands were found across breeds, consistent with possible signatures of either artificial or natural selection. For most genes associated with ROH islands, a biological link to traits of economic importance, which are known to be under selection, can be hypothesized and are consistent with the phenotypic characteristics of these breeds. Because genomic regions that are subjected to selection tend to generate ROH islands, their distributions can indicate genomic regions that may have been subjected to selective pressure. Our results contributed to understanding how selection can shape the distribution of ROH islands and suggested that ROH islands can be used to identify genes potentially involved in economically important traits. Further research must be performed to compare selection signatures and ROH islands, and to incorporate the use of ROH island' distributions across the genome to limit the number of false positives identified and to modify current procedures.

Acknowledgments

This research was financed by PON02_00451_3133441, CUP: B61C1200076005 funded by MIUR.

Declaration of interest

None.

Ethics statement

None.

Software and data repository resources

None.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1751731118000629

References

Barrett JC, Fry B, Maller J and Daly MJ 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265.

Cheng HC, Zhang FW, Deng CY, Jiang CD, Xiong YZ, Li FE and Lei MG 2007. NNAT and DIRAS3 genes are paternally expressed in pigs. Genetics Selection Evolution 39, 599–607.

Cohen M, Reichenstein M, Everts-van der Wind A, Heon-Lee J, Shani M, Lewin HA, Weller JI, Ron M and Seroussi E 2014. Cloning and characterization of FAM13A1-a gene near a milk protein QTL on BTA6: evidence for population-wide linkage disequilibrium in Israeli Holsteins. Genomics 84, 374–383.

Connor EE, Sonstegard TS, Kahl S, Bennett GL and Snelling WM 2003. The bovine type I iodothyronine deiodinase (DIO1) gene maps to chromosome 3. Animal Genetics 34, 233–234.

Ferenčaković M, Hamzić E, Gredler B, Solberg TR, Klemetsdal G, Curik I and Sölkner J 2013a. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. Journal of Animal Breeding and Genetics 130, 286–293.

Ferenčaković M, Solkner J and Curik I 2013b. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. Genetics Selection Evolution 45, 42.

Flori L, Fritz S, Jaffrézic F, Boussaha M, Gut I, Heath S, Foulley JL and Gautier M 2009. The genome response to artificial selection: a case study in dairy cattle. PLoS One 4, e6595.

Gaspa G, Marras G, Sorbolini S, Ajmone-Marsan P, Williams JL, Valentini A, Dimauro C and Macciotta NNP 2014. Genome-wide homozygosity in Italian Holstein cattle using HD panel. In Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, 17 to 22 August, Vancouver, BC, Canada.

Gibson J, Morton N and Collins A 2006. Extended tracts of homozygosity in outbred human populations. Human Molecular Genetics 15, 789–795.

Gutiérrez-Gil B, Arranz JJ and Wiener P 2015. An interpretive review of selective sweep studies in Bos taurus cattle populations: identification of unique and shared selection signals across breeds. Frontiers in Genetics 6, 167.

Huang W, Peñagaricano F, Ahmad KR, Lucey JA, Weigel KA and Khatib H 2012. Association between milk protein gene variants and protein composition traits in dairy cattle. Journal of Dairy Science 95, 440–449.

Kim ES, Cole JB, Huson H, Wiggans GR, Van Tassell CP, Crooker BA, Liu G, Da Y and Sonstegard TS 2013. Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. PLoS One 8, e80813.

Kommadath A, Woelders H, Beerda B, Mulder HA, de Wit AA, Veerkamp RF, te Pas MF and Smits MA 2011. Gene expression patterns in four brain areas associate with quantitative measure of estrous behavior in dairy cows. BMC Genomics 12, 200.

Kukučková V, Moravčíková N, Ferenčaković M, Simčič M, Mészáros G, Sölkner S, Trakovická A, Kadlečík O, Curik I and Kasarda R 2017. Genomic characterization of Pinzgau cattle: genetic conservation and breeding perspectives. Conservation Genetics. 18, 893–910.

Li H, Wang Z, Moore SS, Schenkel FS and Stothard P 2010. Genome-wide scan for positional and functional candidate genes affecting milk production traits in Canadian Holstein Cattle. In Proceedings of the 9th World Congress of Genetics Applied to Livestock Production, 1 to 6 August, Leipzig, Germany.

Li D, Xie X, Wang J, Bian Y, Li Q, Gao X and Wang C 2015. MiR-486 regulates lactation and targets the PTEN gene in cow mammary glands. PLoS One 10, e0118284.

Ma L, O'Connell JR, VanRaden PM, Shen B, Padhi A, Sun C, Bickhart DM, Cole JB, Null DJ, Liu GE, Da Y and Wiggans GR 2015. Cattle sex-specific recombination and genetic control from a large pedigree analysis. PLoS Genetics 11, e1005387.

Mastrangelo S, Saura M, Tolone M, Salces-Ortiz J, Di Gerlando R, Bertolini F, Fontanesi L, Sardina MT, Serrano M and Portolano B 2014. The genome-wide structure of two economically important indigenous Sicilian cattle breeds. Journal of Animal Science 92, 4833–4842.

Mastrangelo S, Tolone M, Di Gerlando R, Fontanesi L, Sardina MT and Portolano B 2016. Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. Animal 10, 746–754.

Mastrangelo S, Tolone M, Sardina MT, Sottile G, Sutera AM, Di Gerlando R and Portolano B 2017. Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. Genetics Selection Evolution 49, 84.

Mészáros G, Boison SA, Pérez O'Brien AM, Ferenčaković M, Curik I, Da Silva MV, Utsunomiya YT, Garcia JF and Sölkner J 2015. Genomic analysis for managing small and endangered populations: a case study in Tyrol Grey cattle. Frontiers in Genetics 6, 173.

Metzger J, Karwath M, Tonda R, Beltran S, Águeda L, Gut M, Gut IG and Distl O 2015. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. BMC Genomics 16, 764.

Mi H, Muruganujan A and Thomas PD 2013. PANTHER in: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Research 41, D377–D386.

Moore SG, Pryce JE, Hayes BJ, Chamberlain AJ, Kemper KE, Berry DP, McCabe M, Cormican P, Lonergan P, Fair T and Butler ST 2016. Differentially expressed genes in endometrium and corpus luteum of Holstein cows selected for high and low fertility are enriched for sequence variants associated with fertility. Biology of Reproduction 94, 19.

Nayeri S, Sargolzaei M, Abo-Ismail MK, Miller S, Schenkel F, Moore SS and Stothard P 2016. Genome-wide association study for lactation persistency, female fertility, longevity, and lifetime profit index traits in Holstein dairy cattle. Journal of Dairy Science 100, 1246–1258.

Nothnagel M, Lu T, Kayser M and Krawczak M 2009. Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. Human Molecular Genetics 19, 2927–2935.

Ogorevc J, Kunej T, Razpet A and Dovc P 2009. Database of cattle candidate genes and genetic markers for milk production and mastitis. Animal Genetics 40, 832–851.

Olsen HG, Meuwissen THE, Nilsen H, Svendsen M and Lien S 2008. Fine mapping of quantitative trait loci on bovine chromosome 6 affecting calving difficulty. Journal of Dairy Science 91, 4312–4322.

Peripolli E, Munari DP, Silva MVGB, Lima ALF, Irgang R and Baldi F 2016. Runs of homozygosity: current knowledge and applications in livestock. Animal Genetics 48, 255–271.

Pimentel ECG, Bauersachs S, Tietze M, Simianer H, Tetens J, Thaller G, Reinhardt F, Wolf E and König S 2011. Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. Animal Genetics 42, 251–262.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ and Sham PC 2007. PLINK: a tool set for whole genome association and population-based linkage analyses. The American Journal of Human Genetics 81, 559–575.

Purfield DC, Berry DP, McParland S and Bradley DG 2012. Runs of homozygosity and population history in cattle. BMC Genetics 13, 70.

Purfield DC, McParland S, Wall E and Berry DP 2017. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PLoS One 12, e0176780.

Qanbari S, Gianola D, Hayes B, Schenkel F, Miller S, Moore S, Thaller G and Simianer H 2011. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. BMC Genomics 12, 318.

Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, Nielsen R and Simianer H 2014. Classic selective sweeps revealed by massive sequencing in cattle. PLoS Genetics 10, e1004148.

Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR and Simianer H 2010. A genome-wide scan for signatures of recent selection in Holstein cattle. Animal Genetics 41, 377–389.

Ramey H, Decker J, McKay S, Rolf M, Schnabel R and Taylor J 2013. Detection of selective sweeps in cattle using genome-wide SNP data. BMC Genomics 14, 382.

Rincon G, Islas-Trejo A, Castillo AR, Bauman DE, German BJ and Medrano JF 2012. Polymorphisms in genes in the SREBP1 signalling pathway and SCD are associated with milk fatty acid composition in Holstein cattle. Journal of Dairy Research 79, 66–75.

Sahana G, Guldbrandtsen B, Thomsen B and Lund MS 2013. Confirmation and fine-mapping of clinical mastitis and somatic cell score QTL in Nordic Holstein cattle. Animal Genetics 44, 620–626.

Sölkner J, Ferenčaković M, Karimi Z, Perez O'Brien AM, Mészáros G, Eaglen S, Boison SA and Curik I 2014. Extremely non-uniform: patterns of runs of homozygosity in bovine Populations. In Proceedings, 10th World Congress of Genetics Applied to Livestock Production, 17 to 22 August, Vancouver, BC, Canada.

Szmatola T, Gurgul A, Ropka-Molik K, Jasielczuck I, Zabek T and Bugno-Poniewierska M 2016. Charateristics of runs of homozygosity in selected cattle breeds maintained in Poland. Livestock Science 188, 72–80.

Weikard R, Widmann P, Buitkamp J, Emmerling R and Kuehn C 2012. Revisiting the quantitative trait loci for milk production traits on BTA6. Animal Genetics 43, 318–323.

Zhang Q, Boichard D, Hoeschele I, Ernst C, Eggen A, Murkve B, Pfister-Genskow M, Witte LA, Grignola FE, Uimari P, Thaller G and Bishop MD 1998. Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. Genetics 149, 1959–1973.

Zhang Q, Calus MP, Guldbrandtsen B, Lund MS and Sahana G 2015a. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. BMC Genetics 16, 88.

Zhang Q, Guldbrandtsen B, Bosse M, Lund MS and Sahana G 2015b. Runs of homozygosity and distribution of functional variants in the cattle genome. BMC Genomics 16, 542.

Zheng X, Ju Z, Wang J, Li Q, Huang J, Zhang A, Zhong J and Wang C 2011. Single nucleotide polymorphisms, haplotypes and combined genotypes of LAP3 gene in bovine and their association with milk production traits. Molecular Biology Reports 38, 4053–4061.