Genome-wide association identifies methane production level relation to genetic control of digestive tract development in dairy cows

M. Pszczola, T. Strabel, S. Mucha & E. Sell-Kubiak

The global temperatures are increasing. This increase is partly due to methane (CH$_4$) production from ruminants, including dairy cattle. Recent studies on dairy cattle have revealed the existence of a heritable variation in CH$_4$ production that enables mitigation strategies based on selective breeding. We have exploited the available heritable variation to study the genetic architecture of CH$_4$ production and detected genomic regions affecting CH$_4$ production. Although the detected regions explained only a small proportion of the heritable variance, we showed that potential QTL regions affecting CH$_4$ production were located within QTLs related to feed efficiency, milk-related traits, body size and health status. Five candidate genes were found: CYP51A1 on BTA 4, PPP1R16B on BTA 13, and NTHL1, TSC2, and PKD1 on BTA 25. These candidate genes were involved in a number of metabolic processes that are possibly related to CH$_4$ production. One of the most promising candidate genes (PKD1) was related to the development of the digestive tract. The results indicate that CH$_4$ production is a highly polygenic trait.

The increase in the global temperature has a serious impact on the environment and humans. Some of these consequences may exceed adaptive capacities of some species, lead to water supplies shortage, melt the glaciers and increase sea level as well as trigger extreme climatic events. The estimated global temperature increase in 2010 due to greenhouse gas (GHG) and aerosol emissions was 0.81 °C in relation to the pre-industrial era. The 0.11 °C of this increase was contributed by methane (CH$_4$) emissions from direct livestock emissions. Most of livestock CH$_4$ emissions are caused by ruminants. CH$_4$ emissions from ruminants are mostly due to enteric fermentation.

In ruminants the enteric fermentation is a consequence of a normal digestive process. One of their stomachs, the rumen, is inhabited by rumen microorganisms, enabling digestion of feed that contains high amounts of fiber. One of the by-products of this digestive process converting the feed provided to the ruminants by the microorganisms is CH$_4$.

The CH$_4$ consists of carbon. When carbon is lost from the body it may no longer be used by the animal as a source of energy. Therefore, apart from its environmental impact, CH$_4$ emission in ruminants has also a potential negative impact on the profitability of animal production. Due to those potential consequences of CH$_4$ emissions from ruminants, mitigation strategies are under investigations. Optional strategies range from adjusting the management to nutritional treatments. These strategies may have a high impact on CH$_4$ production. An additional strategy of mitigating CH$_4$ production might involve selective breeding for lower emitters. Such a strategy could be possible in case of the existence of genetic variation in CH$_4$ production and a favorable genetic association of CH$_4$ and traits in the current breeding goals.

Recent genetic studies on dairy cattle revealed that while most of the variation in CH$_4$ production is due to non-genetic factors (i.e. feed, management and other environmental factors), the genetic component (i.e. genetic variance) in CH$_4$ production also exists. However, information on the extent of genetic control over CH$_4$ production and the genetic architecture of the trait is generally scant. For example, Manzanilla-Pech et al. contributed equally. Correspondence and requests for materials should be addressed to M.P. (email: mbee@up.poznan.pl) or T.S. (email: strabel@up.poznan.pl)
performed GWAS on different methane phenotypes in beef cattle and validated the results on dairy cattle, whereas Van Engelen performed GWAS on Holstein cows using phenotypes predicted from milk and breath analyses.

Only lately the technology for measuring CH$_4$ production both on a large scale and on individual animals has become available. Among others, high throughput measuring techniques are based on breath analyses, since approx. 90% of enteric CH$_4$ is released during eructation events and by breathing. Nonetheless, collection of such phenotypic records is still challenging and data sets are limited. This technique is promising as it is non-invasive, based on infra-red analyses of breath samples and measurements can easily be taken during milking or feeding. The most common application of these techniques is in combination with automatic milking systems (AMS) enabling a relatively long measurement period (duration of the milking); additionally, several observations per cow per day may be collected from a large number of animals. This type of measurement set up enables collection of large volumes of data, which is a prerequisite for genetic analyses.

To our knowledge, to date no reports are available on genome-wide association analyses based on direct measurements of daily CH$_4$ production in dairy cattle. Therefore, the objective of this study was to undertake a genome-wide association study using CH$_4$ phenotypes measured by breath analyzers to unravel the genomic regions controlling CH$_4$ production from dairy cattle.

**Results**

**Detected SNPs.** The genetic variance for daily CH$_4$ production was estimated independently for each level of 2nd order Legendre polynomials. As the first parameter explains most of the variation, only SNP detected with it will be presented and discussed in this study. The GWAS performed on daily CH$_4$ production indicated 50 SNPs with BF > 10 associated with CH$_4$ production in dairy cattle (Fig. 1). Those SNPs were located on 18 different BTA (Tables 1 and 2). From detected SNPs, three had a BF above 30, which is defined as “very strong” association. On BTA 1, 4, 9, 13 and 25 analysis in Haploview indicated six potential candidate QTL regions (Fig. 2). For those regions and two single SNP associations on BTA 9 and 20, a total of 130 candidate genes (protein-coding and non-coding RNA) were located with BIOMART (Table 1).

The three SNP detected for raw phenotypes with BF > 30 and six possible candidate QTL regions explained 0.032% of the total genetic variance (Table 1), whereas the remaining SNPs with 10 < BF < 30 explained 0.122% of this variance (Table 2). Overall this gives a very low result of 0.154% of the total genetic variance explained by detected SNPs.

**Bioinformatics analysis of detected regions.** Out of 130 candidate genes for CH$_4$ production, 46 remained for a further GO Term analysis as known and non-ambiguous genes. For possible candidate genes, 428 different GO Terms were described: 82 cellular component terms, 251 biological process terms and 95 molecular function terms. Based on the GO Terms, five candidate genes were selected as the most promising: CYP51A1 on BTA 4, PPP1R16B on BTA 13, and NTHL1, TSC2, and PKD1 on BTA 25 (Table 3).

Based on Cow QTLdb, 52 QTLs involved in production and reproduction traits were selected as potentially playing a role in daily CH$_4$ production in cows. Those QTLs were clustered into five groups: feed efficiency, milk related, body size and health status (see Table 4).

**Discussion**

To our knowledge this is the first GWAS on direct measurements of daily CH$_4$ production performed in dairy cattle. So far one GWAS on direct measurements of CH$_4$ production was performed in beef cattle with validation on dairy cattle and for CH$_4$ intensity. Another GWAS study on dairy cattle used predicted CH$_4$ following the formula proposed by Dijkstra et al. Thus very little is still known on the actual genomic architecture of CH$_4$ production in dairy cattle. Our results provide more insight into the genomic architecture of CH$_4$ production thanks to the identification of genomic regions involved in the control of this trait and revealed genomic relationships between CH$_4$ production and other traits.

Methane emission may be expressed in several ways depending on the aim of a given study. First of all, when the total CH$_4$ emitted by cows is of interest, the CH$_4$ production phenotype expressed in g/d or l/d may be used. When the goal is to minimize the amount of CH$_4$ emitted from the supplied unit of feed (i.e.
where difference between predicted and measured CH$_4$ emission (i.e. residual CH$_4$ emission)\(^{14,26,29}\). For our analyses physiological status and milk production as described in Pszczola \textit{et al.}\(^{13}\) following Madsen CH$_4$ production\(^{24,33–39}\).

PPP1R16B could affect efficient use of feed and in this way control the amount of by-products (including CH$_4$) produced during the process of digestion. The second gene, namely \textit{PPP1R16B} (BTA13: 68,258,627–68,366,080), a protein phosphatase 1 regulatory subunit 16B, is located within the candidate QTL region on BTA 13. For this gene two biological processes involving \textit{PPP1R16B} suggest that this gene could affect the digestive process by controlling the passage of water within the intestine and providing blood vessels to the endothelial cells of the intestine. Being part of such processes, \textit{PPP1R16B} may affect digestive processes and consequently also a number of their by-products, e.g. CH$_4$, being released post feeding.

The three other genes were all located within the largest detected candidate QTL region on BTA 25, comprising of four SNPs. The first of them, \textit{CYP51A1} (BTA4: 9,306,414–9,323,252) located within the region of a candidate QTL on BTA 4, is a member of the cytochrome P450 family 51 subfamily A. Based on GO Terms this gene is involved in two biological processes that could potentially affect CH$_4$ production in dairy cattle. Those GO Terms are the lipid metabolic process and the steroid metabolic process\(^{30,31}\), which are confirmed by \textit{CYP51A1} and its family members being involved in the synthesis of cholesterol, steroids and other lipids\(^{32}\). Lipids (i.e. fatty acids) were previously reported to be related to CH$_4$ production, including several studies that used fatty acids present in milk to predict CH$_4$ production\(^{4,33–39}\). The three other genes were all located within the largest detected candidate QTL region on BTA 25, comprising of four SNPs. The first of the genes, \textit{NTHL1}, nth like DNA glycosylase 1, is located at 1,590,252–1,595,934 bp. Its gene suggests that \textit{NTHL1} may affect digestive processes and consequently also a number of their by-products, e.g. CH$_4$, being released post feeding.

The second of the above-mentioned genes, \textit{TSC2} (BTA25:1,596,730–1,626,967), tuberous sclerosis 2, is the only candidate gene with a GO Term related to a cellular component, in that case lysosome\(^{31,40}\). Moreover, \textit{TSC2} has been very well studied in humans, as its mutation causes tuberous sclerosis and its product is believed to be a tumor suppressor\(^{35}\). In the case of dairy cattle the location of the \textit{TSC2} gene in lysosome, which contains

| Candidate QTL | Allele subs. effect\(^{a}\) | Number of candidate genes | Total genetic var. expl. (%) |
|---------------|-----------------------------|---------------------------|----------------------------|
| 1 BTA89822nors | 0.113 12 0.006 46223040 0.491 14.33 |
| 1 BTA89820nors | 0.193 0.006 46321775 0.488 12.26 |
| 4 Hapmap39581BTA70101 | 0.226 14 0.003 9203380 0.497 12.26 |
| 4 ARSBFGLNSG109843 | 0.251 0.003 9615916 0.430 14.33 |
| 9 BTB00395654 | 0.132 0.003 60102040 0.353 32.29 |
| 9 ARSBFGLNSG36482 | 0.201 9 0.005 64262480 0.365 16.41 |
| 9 BTB0163493 | 0.172 0.005 64291804 0.365 16.41 |
| 9 Hapmap42513BTA33276 | 0.183 7 0.005 66997852 0.215 23.76 |
| 9 Hapmap27624RAT154889 | 0.158 0.005 67122449 0.318 13.29 |
| 13 BPII | 0.144 21 0.003 67833218 0.402 11.22 |
| 13 ARSBFGLNSG103635 | 0.152 0.003 67888763 0.467 48.68 |
| 20 ARSBFGLNSG109784 | 0.129 1 0.003 909076 0.407 36.61 |
| 25 ARSBFGLNSG61709 | 0.277 0.004 1086505 0.432 17.45 |
| 25 ARSBFGLNSG103099 | 0.238 0.004 1127441 0.395 15.37 |
| 25 ARSBFGLBAC3143 | 0.101 0.003 1184038 0.395 24.82 |
| 25 Hapmap29768RBC016149 | 0.321 0.003 1205232 0.346 16.41 |

Table 1. Candidate QTL regions and single SNPs detected for methane production with Bayesian Factor (BF) above 30, their position in base pairs, minor allele frequency (MAF), number of candidate genes and percentage of total genetic variance explained by them. \(^{a}\)Allele substitution effects were estimated as $\alpha = \sqrt{\sigma^2 q (2pq)}$\(^{-1}\), where $\sigma^2$ is the genetic variance explained by the SNP, and $p$ and $q$ are the frequencies of the two alleles\(^{26,29}\).
hydrolytic enzymes and takes part in energy metabolism, suggest that it could be involved in digestion processes and degradation of metabolites, this may affect CH4 produced by a cow.

The last of the candidate genes on BTA 25 is PKD1, encoding polycystein 1, a transient receptor potentially involved in channel interacting (1,627,978-1,666,088 bp). Three biological processes were assigned to it in GO Term analysis, i.e. blood vessel development 40, nitrogen compound metabolic process 31,40,42,43 and the digestive tract development 40. All three GO Terms indicate that PKD1 is involved in digestion processes either directly by affecting the development of the digestive tract, or possibly also blood vessels around it as well as metabolic processes of a nitrogen compound. All those functions, in general, indicate that PKD1 might be involved in emissions of greenhouse gases, not only CH4 but also nitrogen related.

To confirm that the candidate genes detected in this study are the actual causative mutations affecting CH4 further functional studies such as gene expression or sequencing of the region of highest interest are required. However, this was outside the scope of this paper.

### Potential Quantitative Trait Loci

Next to the search for candidate genes, we have also looked for previously detected QTLs for traits potentially related to CH4 production. Those QTLs were clustered in four groups of similar traits: feed efficiency, milk related, body size and health status (see Table 4). It has to be noted that in this study the estimation of CH4 production included an equation, in which fat-protein-corrected milk, live weight and pregnancy status are taken into account, and some of the found relationships may be present due to this fact. Alternatively, CH4 concentration (expressed in ppm) could be used for the association study. At this moment,
however, CH₄ production is the most widely reported trait in genetic studies regarding reduction of enteric CH₄ emissions. For this reason we restricted our study to this trait.

Firstly, the comparison indicated an overlap between the genomic regions controlling the CH₄ production and QTLs for feed efficiency traits (e.g. residual feed intake, feed conversion ratio, average daily gain; Table 4). The relationship between diet composition and CH₄ production⁴⁴ or the effect of additives reducing emission⁴⁵–⁴⁹ or dry matter intake⁵⁰–⁵² is well known. It is anticipated that increased CH₄ production leads to the loss of energy provided with feed⁵,⁶, and therefore more efficient cows should produce less CH₄. Jentsch et al.⁵³ showed that greater feed ingestion results in higher total CH₄ production; however, CH₄ production per kg dry matter intake decreases. Pickering et al.⁵⁴ also reported the presence of a correlation between CH₄ production and intake, while studies of⁵⁵–⁵⁷ showed that selection for cows with a low residual intake (efficient ones) results in lower CH₄ production. Unfortunately, in this study no data was available on individual feed intake of cows and therefore we were not able to verify this statement empirically.

Figure 2. Results of the linkage disequilibrium (LD) analysis for significant SNPs detected on Bos Taurus autosomes (BTA) for raw phenotypic methane production. (A) BTA 1, (B) BTA 4, (C) BTA 9, (D) BTA 13, (E) BTA 25. Each square contains a value for r² between neighboring SNP.
Secondly, regions controlling CH₄ production were also overlapping with QTLs for traits describing various aspects of milk production (e.g. milk yield, milk protein and fat yield, milk composition; Table 4). The relationship between milk composition and CH₄ production is particularly plausible because of common biochemical pathways between CH₄, acetate and butyrate58. Furthermore, earlier studies showed that it is possible to use milk fatty acid composition to predict CH₄ production24,33,35–39.

Thirdly, it was found that height, chest depth and body weight of the cow were genetically controlled by the same regions as potential QTLs for the CH₄ production. Body characteristics such as body weight were earlier shown to be related to CH₄ production52,59,60. Heavier cows are usually bigger and have a larger rumen capacity and a lower passage rate61, which leads to greater CH₄ production52.

Finally, the QTLs detected previously for the health status of the cow (e.g. mastitis, somatic cell score, immunoglobulin G level) were also found in regions overlapping with SNPs detected in this study for CH₄ production. Thus reports on the relationship between the health status of the animal and the direct CH₄ production are limited. Zetouni et al. (2008) showed a negative genetic correlation on the health of the cows and methane

| Gene       | BTA | Position       | Type of GO Term | GO Term name                                                                 |
|------------|-----|----------------|-----------------|------------------------------------------------------------------------------|
| CYP51A1    | 4   | 9306414–9323252| Biological process| lipid metabolic process, steroid metabolic process                           |
| PPP1R16B   | 13  | 68258627–68366080| Biological process| establishment of an endothelial barrier, positive regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis |
| NTHL1      | 25  | 1590252–1595934 | Biological process| metabolic process                                                             |
| TSC2       | 25  | 1596730–1626967 | Cellular component| Lysosome                                                                      |
| PKD1       | 25  | 1627978–1666088 | Biological process| blood vessel development, nitrogen compound metabolic process, digestive tract development |

Table 3. GO Terms for most promising candidate genes detected for methane production in dairy cattle.

| Group of traits | Trait                          | BTA       |
|-----------------|--------------------------------|-----------|
| Feed efficiency | Residual feed intake           | 4         |
|                 | Feed conversion ratio          | 4         |
|                 | Average daily gain             | 4         |
| Body size       | Height (mature)                | 4         |
|                 | Chest depth                    | 9         |
|                 | Body weight (mature)           | 9         |
| Milk            | Milk fat yield                 | 9; 25     |
|                 | Milk protein yield             | 1; 13; 20 |
|                 | Milk yield                     | 13        |
|                 | Milk energy yield              | 9         |
|                 | cis-Vaccenic acid content      | 20        |
|                 | Docosatetraenoic acid content  | 9         |
|                 | Eicosapentaenoic acid content  | 9         |
|                 | Linoleic acid content          | 1         |
|                 | Milk alpha-casein percentage   | 1         |
|                 | Milk capric acid percentage    | 13        |
|                 | Milk caproic acid percentage   | 13        |
|                 | Milk caprylic acid percentage  | 13        |
|                 | Milk myristoleic acid percentage| 13        |
|                 | Milk palmitoleic acid percentage| 1; 13     |
|                 | Oleic acid content             | 1         |
|                 | Polyunsaturated fatty acid content| 1         |
| Health status   | Somatic cell score             | 4; 9      |
|                 | Clinical mastitis              | 9         |
|                 | Immunoglobulin G level         | 4; 20     |
|                 | Infectious bovine keratoconjunctivitis susceptibility | 1; 20 |
|                 | M. paratuberculosis susceptibility| 20        |

Table 4. Previously detected QTLs within the identified genomic regions potentially related to methane production.
production and a very low positive genetic correlation with udder health. Elliot-Martin et al. (1997), based on breath analyses, indicated that \( \text{CH}_4 \) could be used to diagnose ketosis. Moreover, the health status of the animal is known to affect other traits such as dry matter intake or production, and therefore is likely to affect \( \text{CH}_4 \) production. It is likely that a sick animal produces less methane due to a lower milk production; however, methane intensity (i.e. the amount of methane produced per kg of milk) would increase. Next to QTLs related to traits indicating the health status of the cow also QTLs indicating susceptibility to illness were found in the regions important to \( \text{CH}_4 \) production.

Based on the several traits mentioned above that share the genetic background with \( \text{CH}_4 \) production, it may be suggested that some of the detected regions in this study have a pleiotropic effect. This knowledge is very beneficial especially in the case of production traits controlled by the same regions as \( \text{CH}_4 \) production (i.e. assumed to be genetically correlated), which could serve as indicator traits for enteric \( \text{CH}_4 \) production and eliminate difficult and time-consuming phenotyping. Our findings mostly match the study of Negussie et al., who reviewed literature on potential indirect traits for measuring \( \text{CH}_4 \) production. Further evaluation of genetic relationships between \( \text{CH}_4 \) and other traits is necessary to confirm relationships revealed by our study and before inclusion of \( \text{CH}_4 \) to the breeding program can be made.

**Power of the experimental design.** The Bayesian method selected to perform GWAS for \( \text{CH}_4 \) production allows for good distinctions between SNP with large and small effects on a trait, as in each iteration a different combination of SNPs is given a large effect. Thus detected SNPs give a valuable indication for the genomic regions potentially involved in \( \text{CH}_4 \) production in dairy cattle. This was confirmed also by bioinformatics post-analysis of detected regions with the functions of selected candidate genes and QTLs for other traits detected within those regions. However, the total genetic variance explained by significant SNPs was very low. This could be due to several possible reasons, i.e. (1) a low number of animals used in the study, (2) the accuracy of the collected phenotypes, and (3) the polygenic nature of the studied trait.

Firstly, it should be noted that the analyzed dataset was relatively small, and therefore the power of the GWAS design was too low to detect a majority of SNPs associated with \( \text{CH}_4 \) production. Taking into account the heritability of this trait at 0.27, a higher number of genotyped animals would be needed to obtain a higher percentage of genetic variance explained by the detected SNP. Therefore, the analyses of a larger dataset (for both phenotypic and genomic data) may shed light on more specific SNPs with large effects. However, generating a large data set by one project is difficult due to related costs (measuring and genotyping). Therefore, generating such a dataset by combining phenotypic observations and genotypes from various experiments could be a solution producing more reliable results in the future.

Secondly, to obtain reliable GWAS results reliable phenotypes are needed. In our study we used a technique that measures \( \text{CH}_4 \) at the AMS during milking. To verify the accuracy of the sensor used in this study we validated the used sensor against sensors used in Respiration Chambers (the standard \( \text{CH}_4 \) measuring technique). This comparison showed a high similarity between results generated by the two sensors when used in the AMS. There are no studies comparing the performance of sensors used in the present study when installed in the Respiration Chamber. Several factors could lead to inaccuracies in the collected measurements such as occasional wind in the area of AMS or cows’ head movement. These factors were not controlled in this study. To account for these arguably random effects we measured \( \text{CH}_4 \) for the individuals in the long period of time (i.e. resulting in multiple observations per animal). The average repeatability of the analyzed phenotype was 0.25 as reported in Pszczola et al.

Thirdly, the greater data set and increased accuracy of the measuring method could not have been enough to explain more genetic variation if the analyzed trait was highly polygenic. In previous studies using the same methodology, but larger data sets, only 0.83% of the genetic variance was explained by SNP in GWAS on litter size in pigs and 9.5% in GWAS on teat number in pigs. Based on the presented results it seems that \( \text{CH}_4 \) production is also a highly polygenic trait and many different regions are involved in its regulation. It might not be, therefore, possible to detect all of them using GWAS.

As \( \text{CH}_4 \) production turned out to be a very polygenic trait in application to breeding practice, it may be more advisable to use the genomic prediction approach without specifying particular SNPs as being more important than others (e.g. genomic BLUP). In fact, de Haas et al., Lassen et al., and Wilson et al. performed genomic prediction type analyses while searching for correlated traits. The biggest challenge for the performance of genomic prediction with sufficient, reasonable or high accuracy of the estimated genotypic values is to create an adequately large reference population, which is likely to require cooperation between several countries.

**Conclusions**

This study aimed at detecting genomic regions affecting \( \text{CH}_4 \) production in dairy cattle and showed that SNPs associated with the trait of interest may be detected. However, \( \text{CH}_4 \) data collection poses a challenge, leading to a lower power of the experimental design and prevented detection of a high number of SNPs with a large effect on \( \text{CH}_4 \) production. Consequently, only a small proportion of the genetic variance was explained by the SNPs. Nonetheless, the candidate QTL region on BTA 25, where three candidate genes were identified, may be considered as a genomic region regulating \( \text{CH}_4 \) production in dairy cattle. Furthermore, the comparison of the QTL regions affecting \( \text{CH}_4 \) production with previously reported QTLs indicated common genomic regions between \( \text{CH}_4 \) production and traits related to feed efficiency, milk related, body size and health status. The found candidate genes were also involved in a number of metabolic processes possibly related to \( \text{CH}_4 \) production. One of the most promising candidate genes (PKD1) was related to the development of the digestive tract being the environment inhabited by methanogens and the site for methane production. In general, all the evidence shows that \( \text{CH}_4 \) production is a polygenic trait.
Methods
All research was approved by the Local Ethical Committee for Experiments on Animals in Poznan, Poland (Decision Number: 64/2012) and performed in accordance with the “Act on the protection of animals used for scientific purpose” of the Republic of Poland, which complies with the European Union Legislation for the protection of animals used for scientific purposes.

Phenotypes. The observations on CH\(_4\) production [g/d] used in this study were obtained from Pszczola et al\(^{13}\), where all the detailed information on farms, measuring set-up and data processing can be found.

In short, animals available for this study were 287 Polish Holstein-Friesian cows kept on two commercial farms in Poland. This was a subset of 483 cows genotyped for CH\(_4\) production and analyzed by Pszczola et al\(^{13}\), of which 287 were genotyped. The CH\(_4\) production was measured repeatedly on Farm1 during two periods: from 2014/12/02 to 2016/02/03, and from 2016/06/01 to 2016/09/17, and on Farm2 from 2016/02/05 to 2016/03/14. Cows were milked repeatedly during the experiment, in total 25,872 CH\(_4\) production observations were collected for the genotyped animals.

The CH\(_4\) production was measured using a non-invasive Fourier Transform Infrared Spectroscopy breath analyzer (GASMET 4030; Gasmet Technologies Oy, Helsinki, Finland) during milking in AMS (Lely Astronaut A4). Concentrations of CH\(_4\) and CO\(_2\) measured during milking were converted to daily CH\(_4\) production in grams per day [g/d] following Madsen et al\(^{28}\) and Pedersen et al\(^{32}\). This calculation took into account the concentrations of CH\(_4\) and CO\(_2\), fat-protein corrected milk, live weight and duration of the pregnancy. Multiple daily outputs per cow were corrected for the diurnal variation in CH\(_4\) and averaged per cow per day.

Genotypes. Cows were genotyped with the Illumina BovineSNP50 v2.0 BeadChip (Illumina Inc., San Diego, CA) at the Cattle Genetics Laboratory of the Polish Federation of Cattle Breeders and Dairy Farmers. Ear tissue samples used to extract DNA were collected in the course of a routine procedure within the breeding program. The genotyped SNPs were processed with following quality control checks: (1) being in Hardy–Weinberg equilibrium, (2) having the minor allele frequency above 0.05, (3) not being monomorphic, and (4) having a call rate of above 0.95. Six cows were removed as they had the call rate below 0.9. After quality control and removing SNPs located on sex chromosomes and chromosome 0 (unassigned), 39,680 SNPs remained for the genome-wide association analysis.

Genome-wide association. To identify regions of the genome affecting CH\(_4\) production, a multi-SNP genome-wide association analysis was performed with the application of the Bayesian Variable Selection method\(^{31}\). The method allows for a simultaneous estimation of the effects of all markers used in the analysis. The analysis was performed with the Bayz software\(^{22}\) on daily CH\(_4\) production using the model developed by Pszczola et al\(^{13}\). The model was:

\[
\text{CH}_4 = \mu + \mathbf{Xb} + \sum_{n=1}^{3} \mathbf{DIM}_{ij} + \epsilon \\
\text{with } \mathbf{L}_k \times \sum_{n=1}^{3} \mathbf{DIM}_{ij} = \mathbf{Z}_d \beta_{ik} + \epsilon_{ijk},
\]

where \(\text{CH}_4\) stands for the daily CH\(_4\) production levels of a cow; \(\mu\) is an n-vector equal to the mean; \(\mathbf{Xb}\) is the design matrix of fixed effects of year-week of measurement and cow’s lactation number (levels 1 or 2+) fitted within the general lactation curve, which was modeled using 3\(^{rd}\) order Legendre polynomials; and \(\epsilon\) is an n-vector of random residual effects assumed to be normally distributed \(N(0, \sigma^2_\epsilon)\). The \(\mathbf{L}_k\) is a vector of individual random animal effect, which was modeled using 2\(^{nd}\) order Legendre polynomials. The mapping of marker effects is constructed as a hierarchical model on random animal effects\(^{73}\). Firstly, the model accounts for genetic variance only. Secondly, at the next level the model allows disentangling permanent environmental (Note: this accounted for repeated observations of daily CH\(_4\) production per cow.) and genetic variances independently for each level of 2\(^{nd}\) order Legendre polynomials. Here the \(\mathbf{Z}_d\) is a matrix with dimensions \(n \times p\), with \(n\) being the number of genotypes and \(p\) being the number of SNP coded as 0, 1, 2 copies of a specific allele vector; \(\beta_{ik}\) is a p-vector with the random effects of markers; and \(\epsilon_{ijk}\) accounts for the permanent environmental effect assumed to be normally distributed \(N(0, \sigma^2_{\epsilon_{ijk}})\).

For the marker effect the Bernoulli distribution was applied:

\[
\beta \sim \begin{cases} 
N(0, \sigma^2_\beta) & \text{with probability: } \tau_0 \\
N(0, \sigma^2_\beta) & \text{with probability: } \tau_1
\end{cases}
\]

where for the first distribution it is assumed that the SNPs have a small effect (\(\sigma^2_\beta\)); whereas in the second distribution the SNPs are assumed to have a large effect, which explains a large part of variance (\(\sigma^2_\beta\)) of analyzed traits. In this study, a prior of \(\tau_1 = 0.001\) was selected, thus on average only 1 in 1,000 SNPs was in the second distribution in each cycle. This resulted in only ~38 SNPs per cycle to have a large effect on the traits. The posterior means were calculated with 500k MCMC iterations with burn-in of 5k iterations to secure that all the SNPs were used\(^{69,66,74}\). Selecting a stringent prior provides a more precise distinction between SNPs with large and small effects on the trait\(^{66,25}\). If the SNP was not genotyped for a certain animal then Bayz assigned an average genotype to that position.
Identification of significant SNPs. The Bayes Factor (BF) was calculated for each SNP to determine the significant associations:

$$BF = \frac{\hat{\pi}_i}{\hat{\pi}_0},$$

where $\pi_i$ and $\pi_0$ are the prior probabilities and $\hat{\pi}_i$ is the posterior probability of the fraction of times the SNP was in the distribution with a large effect. Following the definitions of Kass and Raftery\(^{20}\), the SNPs with BF > 30 are described as a "very strong" association and with BF > 150 as "decisive". The variance explained by significant SNPs was estimated as a fraction of the total genetic variance explained by all SNPs.

To confirm the potential QTL regions, also the linkage disequilibrium (LD) measured by $r^2$ was estimated in Haploview\(^{21}\) between the SNPs detected on one BTA and not further from each other than 500 kbp. The candidate gene search was performed with the BIOMART software available in Ensembl Bos Taurus UMD 3.1\(^{18}\) by entering the position of a possible QTL region or one of the most significant SNPs with ±500 kbp. To limit the number of QTLs to the most promising as candidate genes for daily CH\(_4\) production the BIOMART database was also used to study Gene Ontology Terms (GO Terms) of those QTLs. Furthermore, the Cow QTL database of the Animal Genome project\(^{20}\) was used to find previously detected QTLs within the most promising regions detected here for daily CH\(_4\) production. This was done analogically as for the candidate gene search, i.e. by entering the position of a possible QTL region or one of the most significant SNPs with ±500 kbp.

References

1. Stern, N. Stern Review Report on the Economics of Climate Change (2006).
2. Reisinger, A. & Clark, H. How much do direct livestock emissions actually contribute to global warming? Global change biology 24, 1749–1761 (2018).
3. Opio, C. et al. Greenhouse gas emissions from ruminant supply chains – A global life cycle assessment. Food and Agriculture Organization of the United Nations (FAO), Rome, 1–214 (2013).
4. Smith, P. et al. Agriculture, forestry and other land use (AFOLU) (2014).
5. Murray, R. M., Bryant, A. M. & Leng, R. A. Rates of production of methane in rumin and large-intestine of sheep. Be. J. Nutr. 36, 1–14, https://doi.org/10.1079/bjn19760053 (1976).
6. Johnson, D. E. et al. In Atmospheric methane: sources, sinks, and role in global change Vol. 13 (ed Khalil, M.A.K.) Ch. 11, 199–229 (Springer, 1993).
7. Knapp, J. R., Laur, G. L., Vadas, P. A., Weiss, W. P. & Tricarico, J. M. Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. J. Dairy Sci. 97, 3231–3261, https://doi.org/10.3168/jds.2013-7234 (2014).
8. Lassen, J. & Lovendahl, P. Heritability estimates for enteric methane emissions from Holstein cattle measured using noninvasive methods. J Dairy Sci. 99, 1959–1967, https://doi.org/10.3168/jds.2015-10012 (2016).
9. Lassen, J. & Lovendahl, P. Heritability for enteric methane emission from Danish Holstein cows using a non-invasive FTIR method. Adv. Anim. Biosci. 4, 280 (2013).
10. de Haas, Y. et al. Genetic parameters for predicted methane production and potential for reducing enteric emissions through genomic selection. J. Dairy Sci. 94, 6122–6134, https://doi.org/10.3168/jds.2011-4439 (2011).
11. Yin, T., Pinent, T., Brügemann, K., Simianer, H. & König, S. Simulation, prediction, and genetic analyses of daily methane emissions in dairy cattle. J. Dairy Sci. 98, 5748–5762, https://doi.org/10.3168/jds.2014-6818 (2015).
12. van Engelen, S., Bovenhuis, H., Dijkstra, J., van Arendonk, J. & Visker, M. Short communication: Genetic study of methane production predicted from milk fat composition in dairy cows. J. Dairy Sci. 98, 8223–8226, https://doi.org/10.3168/jds.2014-8989 (2015).
13. Pszczola, M., Rzewuska, K., Mucha, S. & Strabel, T. Heritability of methane emissions from dairy cows over a lactation measured on commercial farms. Journal of Animal Science 100, 2231–2237, https://doi.org/10.1093/jas/sky202 (2012).
14. Macanilla-Pech, C. I. V. et al. Genomewide association study of methane emissions in Angus beef cattle with validation in dairy cattle. Journal of Animal Science 94, 4151–4166, https://doi.org/10.2527/jas.2016-0431 (2016).
15. Engelen, S. V. The genetic background of methane emission by dairy cows, Wageningen University (2018).
16. Lassen, J., Lovendahl, P. & Madsen, J. Accuracy of noninvasive breath methane measurements using Fourier transform infrared methods on individual cows. J. Dairy Sci. 95, 890–898, https://doi.org/10.3168/jds.2011-4544 (2012).
17. Garnsworthy, P. C., Craigon, J., Hernandez-Medrano, J. H. & Saunders, N. On-farm methane measurements during milking correlate with total methane production by individual dairy cows. J. Dairy Sci. 95, 3166–3180, https://doi.org/10.3168/jds.2011-4605 (2012).
18. Negussie, E. et al. Non-invasive individual methane measurement in dairy cows. animal 11, 890–899, https://doi.org/10.1017/ s1751731116002718 (2016).
19. Chagunda, M. G. G., Ross, D. & Roberts, D. J. On the use of a laser methane detector in dairy cows. Computers and Electronics in Agriculture 68, 157–160, https://doi.org/10.1016/j.compag.2009.05.008 (2009).
20. Kass, R. E. & Raftery, A. E. Bayes Factors. Journal of the American Statistical Association 90, 773–795, https://doi.org/10.1080/01621459.1995.10476572 (1995).
21. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265, https://doi.org/10.1093/bioinformatics/bth457 (2005).
22. Cow QTL data base. http://www.animalgenome.org/cgi-bin/QTLdb/BT/index (2017).
23. Van Engelen, S., Bovenhuis, H., Dijkstra, J., van Arendonk, J.A.M. & Visker, M. H. P. Genome wide association studies for milk fatty acids as a basis for methane prediction. Proceedings, 4–6 (2005).
24. Dijkstra, J. et al. Relationships between methane production and milk fatty acid profiles in dairy cattle. Animal Feed Science and Technology 166-67, 590–595, https://doi.org/10.1016/j.anifeedsci.2011.04.042 (2011).
25. Alcock, D. I. & Hegarty, R. S. Potential effects of animal management and genetic improvement on enteric methane emissions, emissions intensity and productivity of sheep enterprises at Cowra, Australia. Animal Feed Science and Technology 166-167, 749–760, https://doi.org/10.1016/j.anifeedsci.2011.04.053 (2011).
26. Herd, R. M. et al. Measures of methane production and their phenotypic relationships with dry matter intake, growth, and body composition traits in beef cattle 1,2. Journal of Animal Science 92, 5262–5274, https://doi.org/10.2527/jas.2014-8273 (2014).
27. de Haas, Y., Pszczola, M., Soyeurt, H., Wall, E. & Lassen, J. Invited review: Phenotypes to genetically reduce greenhouse gas emissions in dairying. J. Dairy Sci., https://doi.org/10.3168/jds.2016-11246 (2017).
28. Madsen, J., Bjerg, B. S., Hvelplund, T., Weisbjerg, M. R. & Lund, P. Methane and carbon dioxide ratio in excreted air for quantification of the methane production from ruminants. Livestock Science 129, 223–227, https://doi.org/10.1016/j.livsci.2010.01.001 (2010).
68. Lassen, J., Poulsen, N. A., Larsen, M. K. & Buitenhuis, A. J. Genetic and genomic relationship between methane production measured in breath and fatty acid content in milk samples from Danish Holsteins. *Animal Production Science* **56**, 298–303, https://doi.org/10.1071/AN15489 (2016).

69. Wilson, A. M. et al. An international effort to improve feed efficiency and reduce methane emissions in dairy cows through genomics. *Journal of Animal Science* **94**, 152, https://doi.org/10.2527/jam2016-0318 (2016).

70. Pedersen, S. et al. Carbon Dioxide Production in Animal Houses: A literature review. *Agricultural Engineering International X* (2008).

71. George, E. I. & McCulloch, R. E. Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association* **88**, 881–889, https://doi.org/10.1080/01621459.1993.10476353 (1993).

72. Heuven, H. C. M. & Janss, L. G. Bayesian multi-QTL mapping for growth curve parameters. *BMC Proceedings* **4**, S12, https://doi.org/10.1186/1753-6561-4-s1-s12 (2010).

73. Heuven, H., GJM., R & L. Janss. In 10th World Congress on Genetics Applied to Livestock Production (WCGALP) 2011–2013 (2014).

74. Lopes, M. S. et al. Using markers with large effect in genetic and genomic predictions. *Journal of Animal Science* **95**, 59–71 (2017).

75. Van Den Berg, I., Fritz, S. & Boichard, D. QTL fine mapping with Bayes C (pi): a simulation study. *Genet. Sel. Evol.* **45**(10), 1186 (2013).

76. Weller, J. I. Quantitative trait loci analysis in animals (2009).

Acknowledgements
The authors would like to thank Luc Janss (Aarhus University) for valuable comments regarding the application of Bayz to dairy cattle data. This project was financed by the Polish National Center for Science (NCN OPUS grant no. 2013/09/B/NZ9/03179). MP and ESK acknowledge the financial support of the Polish Ministry of Science and Higher Education (grant no. 666/2014 and 1021/STYP/12/2017). Part of conducted research and publication of this manuscript was made possible by the statutory funding No. 508.534.01.6 of the Faculty of Veterinary Medicine and Animal Science Poznan University of Life Sciences, Poland; Department of Genetics and Animal Breeding. This study is partly based on the knowledge gained from the networks of COST Action FA1302 “Large-scale methane measurements on individual ruminants for genetic evaluations”.

Author Contributions
M.P. and E.S.K. designed the study and wrote the manuscript. M.P. was responsible for the collection of the phenotypic data, data editing, performing the statistical analysis of phenotypes and part of the post-genome-wide association study analysis. E.S.K. performed the genome-wide association and post analysis of those results. S.M. was responsible for the editing of the genotypes and was a discussion partner with respect to genome-wide association analysis. T.S. took part in data collection and was a discussion partner with respect to the statistical analysis of the phenotypic data and genome-wide association analysis. T.S. and M.P. initialized the project. All authors have read and approved the final manuscript.

Additional Information
Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2018