

METHODOLOGY ARTICLE

Open Access

# A transcriptome map of perennial ryegrass (*Lolium perenne* L.)

Bruno Studer<sup>1\*</sup>, Stephen Byrne<sup>1</sup>, Rasmus O. Nielsen<sup>2</sup>, Frank Panitz<sup>2</sup>, Christian Bendixen<sup>2</sup>, Md. Shofiquel Islam<sup>1</sup>, Matthias Pfeifer<sup>3</sup>, Thomas Lübberstedt<sup>4</sup> and Torben Asp<sup>1</sup>

## Abstract

**Background:** Single nucleotide polymorphisms (SNPs) are increasingly becoming the DNA marker system of choice due to their prevalence in the genome and their ability to be used in highly multiplexed genotyping assays. Although needed in high numbers for genome-wide marker profiles and genomics-assisted breeding, a surprisingly low number of validated SNPs are currently available for perennial ryegrass.

**Results:** A perennial ryegrass unigene set representing 9,399 genes was used as a reference for the assembly of 802,156 high quality reads generated by 454 transcriptome sequencing and for *in silico* SNP discovery. Out of more than 15,433 SNPs in 1,778 unigenes fulfilling highly stringent assembly and detection parameters, a total of 768 SNP markers were selected for GoldenGate genotyping in 184 individuals of the perennial ryegrass mapping population VrnA, a population being previously evaluated for important agronomic traits. A total of 592 (77%) of the SNPs tested were successfully called with a cluster separation above 0.9. Of these, 509 (86%) genic SNP markers segregated in the VrnA mapping population, out of which 495 were assigned to map positions. The genetic linkage map presented here comprises a total of 838 DNA markers (767 gene-derived markers) and spans 750 centi Morgan (cM) with an average marker interval distance of less than 0.9 cM. Moreover, it locates 732 expressed genes involved in a broad range of molecular functions of different biological processes in the perennial ryegrass genome.

**Conclusions:** Here, we present an efficient approach of using next generation sequencing (NGS) data for SNP discovery, and the successful design of a 768-plex Illumina GoldenGate genotyping assay in a complex genome. The ryegrass SNPs along with the corresponding transcribed sequences represent a milestone in the establishment of genetic and genomics resources available for this species and constitute a further step towards molecular breeding strategies. Moreover, the high density genetic linkage map predominantly based on gene-associated DNA markers provides an important tool for the assignment of candidate genes to quantitative trait loci (QTL), functional genomics and the integration of genetic and physical maps in perennial ryegrass, one of the most important temperate grassland species.

**Keywords:** Illumina GoldenGate genotyping, *In silico* SNP discovery, Next generation sequencing (NGS), Perennial ryegrass (*Lolium perenne* L.), Single nucleotide polymorphism (SNP), Transcriptome sequencing

## Background

High density genetic linkage maps are important tools for QTL fine mapping, map-based cloning, comparative genome analysis and the integration of genetic and physical maps. Several genetic linkage maps based on various markers technologies are now available for perennial

ryegrass [1-9]. These maps of moderate marker densities have proved valuable for mapping QTL to broad genome regions. Public marker resources recently established provide the opportunity to increase marker density of these maps, thereby improving map resolution [10-13].

For example, the genetic linkage map of the perennial ryegrass mapping population VrnA has initially been used for a QTL study to characterise vernalization response and contained 93 markers spanning 490.4 cM with an average distance between markers of 5 cM [2]. This map has been complemented over time with candidate gene-based CAPS

\* Correspondence: bruno.studer@agrsci.dk

<sup>1</sup>Department of Molecular Biology and Genetics, Faculty of Science and Technology, Research Centre Flakkebjerg, Aarhus University, Forsøgsvej 1, 4200 Slagelse, Denmark

Full list of author information is available at the end of the article

markers to study disease resistance traits [14,15] and contained around 180 markers with total map length of 487 cM when used to evaluate seed yield and fertility traits [16]. Recently, the same map has been used to localise genes involved in water stress and contained 222 markers, between 24 and 37 on each linkage group (LG), spanning a total of 736 cM [17].

Among the different marker technologies available to increase the density of a genetic linkage map, SNPs have attracted much interest, mainly for two reasons: Firstly, SNPs are the most abundant form of genetic variation [18] and occur at regular intervals in the genome [19]. Secondly, SNPs are highly suitable for multiplexed genotyping assays on mass spectrometry, microarray or beadarray-based platforms [20]. Advancements in these technologies has enabled increased throughput at low cost per data point.

The potential of SNPs for extensive genome analysis has been impressively demonstrated in model plant species such as *Arabidopsis thaliana*, rice (*Oryza sativa*), and maize (*Zea mays*), where fully sequenced genomes resulted in the identification of millions of SNPs suitable for genome-wide association studies and molecular breeding concepts such as genomic selection [21].

In species where a reference genome sequence has not been established yet, several strategies for large-scale SNP discovery have been reported, mainly being divided into *in vitro* and *in silico* approaches. Amplicon resequencing is an *in vitro* approach and has proven very reliable for SNP identification with a false discovery rate usually below 5% [22]. Furthermore, cloned PCR fragments and allele-specific sequencing allow haplotype identification at sufficient read lengths and the discrimination of orthologous (allelic) and paralogous (derived from closely related genes or highly conserved domains in gene families) sequences. However, amplicon resequencing requires an enormous effort for large-scale studies, since each gene needs to be amplified individually and thus might have limited application in the future. Despite the labour intensive nature of amplicon cloning and sequencing, this has been the method of choice for SNP discovery in ryegrasses to date [23]. For *in silico* SNP discovery, the rapidly growing public EST databases can be exploited as a potential sequence resource [24,25]. This approach has been applied in other Poaceae crop species including wheat (*Triticum aestivum* L.) [26] and barley (*Hordeum vulgare* L.) [27]. However, availability and quality of public ryegrass EST sequences are often limited and it might be difficult to obtain a sufficient number of EST reads from the same gene, a key factor for reliable *in silico* SNP identification [22,28]. As a result of these limitations, the percentage of false discovery rates is often considerably high and can vary between 5 and 50% [22]. Recent advances in NGS opened up the opportunity for whole genome resequencing as an extremely powerful strategy for *in silico*

SNP discovery at appropriate sequence coverage. However, *de novo* assembly of short NGS reads is difficult in outbreeding species with a highly heterozygous, large and complex genome containing a high degree of repetitive elements. Moreover, whole genome resequencing may not be necessary to target recombination blocks present in bi-parental mapping populations. Therefore, different strategies for complexity reduction such as reduced representation libraries (RRL) have been proposed to sequence only a subset of the genome for SNP discovery [29]. RRLs have been applied in a wide range of plant species such as maize [30], rice [31], grapevine species (*Vitis* spp.) [32], common bean (*Phaseolus vulgaris* L.) [33] and soybean (*Glycine max* L.) [34]. Another strategy for complexity reduction is transcriptome sequencing [35,36], where expressed genes are targeted and highly repetitive non-transcribed genomic regions are excluded. This emerged as an efficient method for the high-throughput acquisition of gene-associated SNPs [37,38].

For SNP genotyping in a scale up to 3,072 SNPs, the Illumina GoldenGate technology [39] has successfully been used in several crop species. In diploid barley, for example, custom oligo pool assays (OPAs) have been designed to estimate linkage disequilibrium (LD) in inbred elite varieties [40] and for genetic linkage mapping [41]. Recently, two validated 1,536-SNP barley OPAs (BOPA1 and BOPA2) were made available to the barley community as an excellent marker resource in terms of distribution and density in the barley genome, technical performance and biological importance [42]. In more complex genomes such as soybean, GoldenGate genotyping has been used for linkage mapping in recombinant inbred line mapping populations [43]. While also being autogamous, soybean contains around twice as many gene paralogues (32%) when compared to 16% in barley [44], which is known to affect the success rate of multiplexed high-throughput genotyping methods [45,46]. However, the rate of 89% successfully scored SNPs indicated that the genome complexity of soybean had limited impact on GoldenGate performance in a carefully selected SNP panel [43]. In maize, the genome contains about 80% repetitive sequences and a similar amount of paralogous sequences as soybean [44], but a substantially higher intraspecific genetic variation [47]. Despite this, OPAs containing 1,536 SNPs designed from publicly available SNPs (www.panzea.org) are routinely used for diversity, linkage and association analysis, as well as for LD estimations [48,49]. To date, the GoldenGate assay proved even successful for SNP genotyping in tetraploid and hexaploid wheat lines [50] and allopolyploid *Brassica napus* [51].

Encouraged by this, we developed the first open access Lolium 768-SNP OPA (thereafter referred to as LOPA1) for the allogamous forage grass species *L. perenne* with a genome size and complexity comparable to maize.

Specifically, we aimed at (i) developing an efficient strategy for *in silico* SNP discovery based on next generation transcriptome sequencing, (ii) implementing a pipeline for successful OPA design, (iii) getting first insights to cross-species amplification rates of ryegrass SNPs and (iv) constructing a high density EST map in perennial ryegrass as a promising tool for QTL fine mapping, map-based cloning and comparative genome analysis.

## Results

### SNP discovery

A comprehensive EST collection consisting of a total of 31,379 ryegrass ESTs generated by Sanger sequencing was subjected to quality filtering and vector clipping, resulting in 25,744 high-quality EST reads of 8.5 Mbp nucleotide information [52]. A *de novo* assembly using the PHRED, PHRAP, and CROSS\_MATCH software packages resulted in 9,399 non-redundant contigs and singletons with an average length of 889 bp, thereafter referred to as unigene set.

For SNP discovery, 454 GS FLX transcriptome sequencing of the parents of VrnA and a ryegrass genotype that has been inbred for six generations was performed. In total, 802,156 high-quality reads with an average read length of 377 bp were aligned against the unigene set. A minimum of four reads at the SNP position and at least two reads for each SNP variant was required for SNP calling. A total of 15,433 SNPs in 1,778 of these unigenes met the stringent SNP calling parameters, out of which one SNP in each unigene was selected for further analysis.

### SNP selection, validation and *Lolium oligo* pool assays (LOPA1) design

Out of a total of 1,778 SNP-containing unigenes, 556 (31%) were discarded because (i) the detected SNPs were located within a distance of 30 bp to the sequence end or intron/exon splice junctions estimated by BLASTN analysis against the rice genome sequence, (ii) additional SNPs and/or InDels were observed within a distance of 30 bp to the target SNP, or (iii) the reference inbred genotype revealed allelic sequence polymorphisms, indicating the presence of similar but non-allelic sequences in the alignment. For another 132 unigenes (7%), no significant ( $E < e^{-10}$ ) sequence similarities to the rice genome sequence were found by BLASTN analysis, making a proper positional prediction of intron/exon splice junctions impossible. Moreover, sequence reads from only one parental genotype were observed for 72 (4%) of the SNP-containing unigenes.

In order to validate the remaining 1,018 SNPs prior to the GoldenGate assay, a subset of 22 randomly selected SNPs were tested either by direct sequencing of PCR fragments amplified from the mapping parent(s) being polymorphic for the respective SNP or by high resolution

melting (HRM) curve analysis of short amplicons covering the predicted SNP polymorphism (Additional file 1: Figure S1A and S1B). As a result, 17 (77%) out of the 22 examined SNP candidates were experimentally confirmed and represented biological SNPs. Sequencing failed for two SNPs and an additional three (14%) were monomorphic. These five SNPs were excluded from further analysis.

The remaining 1,013 SNPs were subjected to functionality score calculation by Illumina Technical Service, out of which 253 (13%) yielded scores lower than 0.6 and were, therefore, discarded. For eight out of 760 unigenes, two SNP markers were selected for genotyping. Finally, 768 SNPs satisfying the stringent selection criteria were used to design the 768-plex LOPA1.

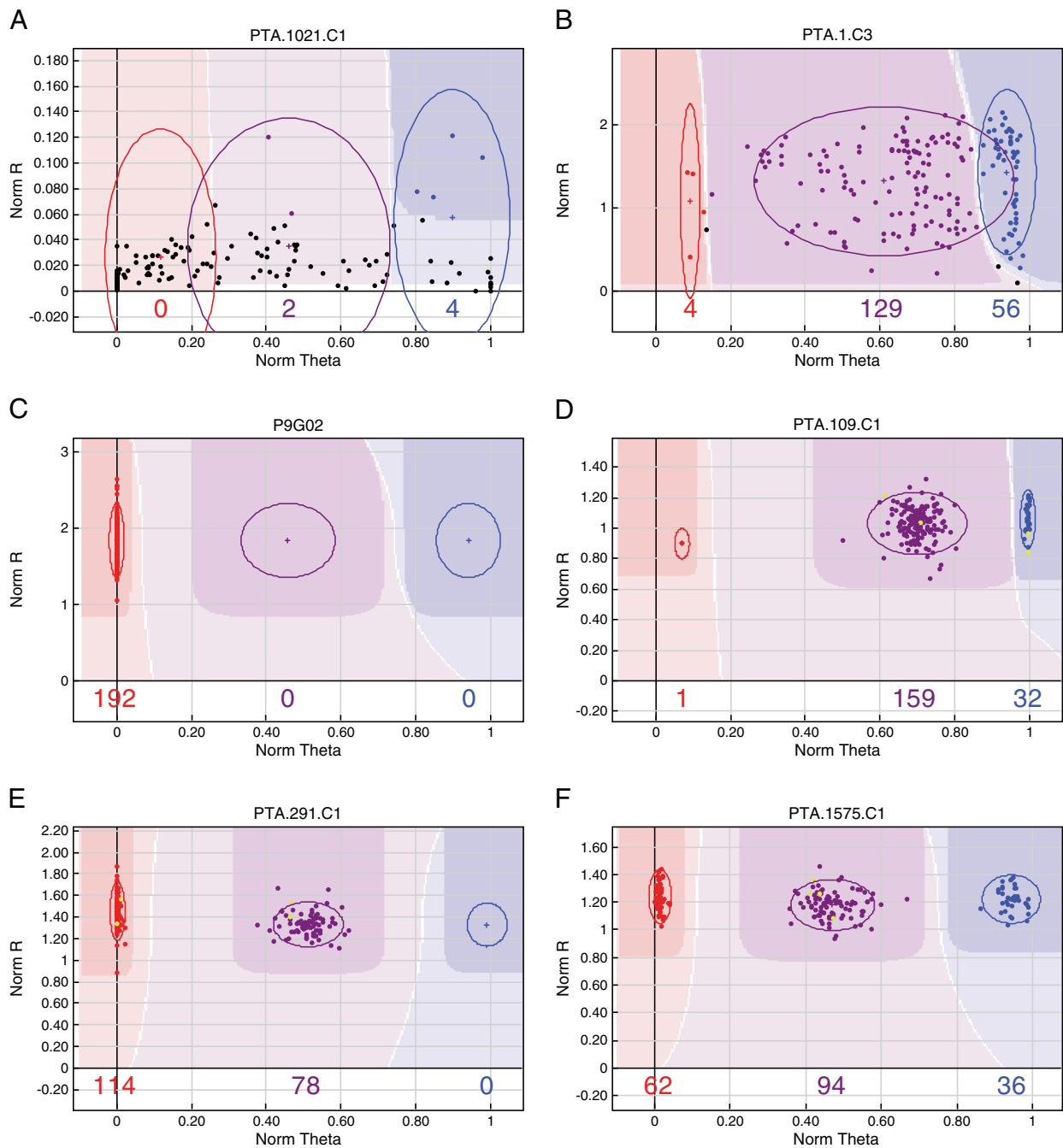
### GoldenGate genotyping and allele calling

The GoldenGate assay failed for 76 out of 768 genotyped SNPs (10%) and poor or inaccurate fluorescent signals were detected (see Figure 1A as an example). Of the remaining 692 SNPs, 100 (14%) did not form clusters reliably separating genotypes and/or revealed cluster separation scores lower than 0.8 (Figure 1B). Additional 83 SNPs (12%) were monomorphic in the mapping population (Figure 1C). The remaining 509 SNPs (77%) were segregating either in one (Figure 1D and 1E) or in both mapping parents (Figure 1F) and were available for genetic linkage mapping.

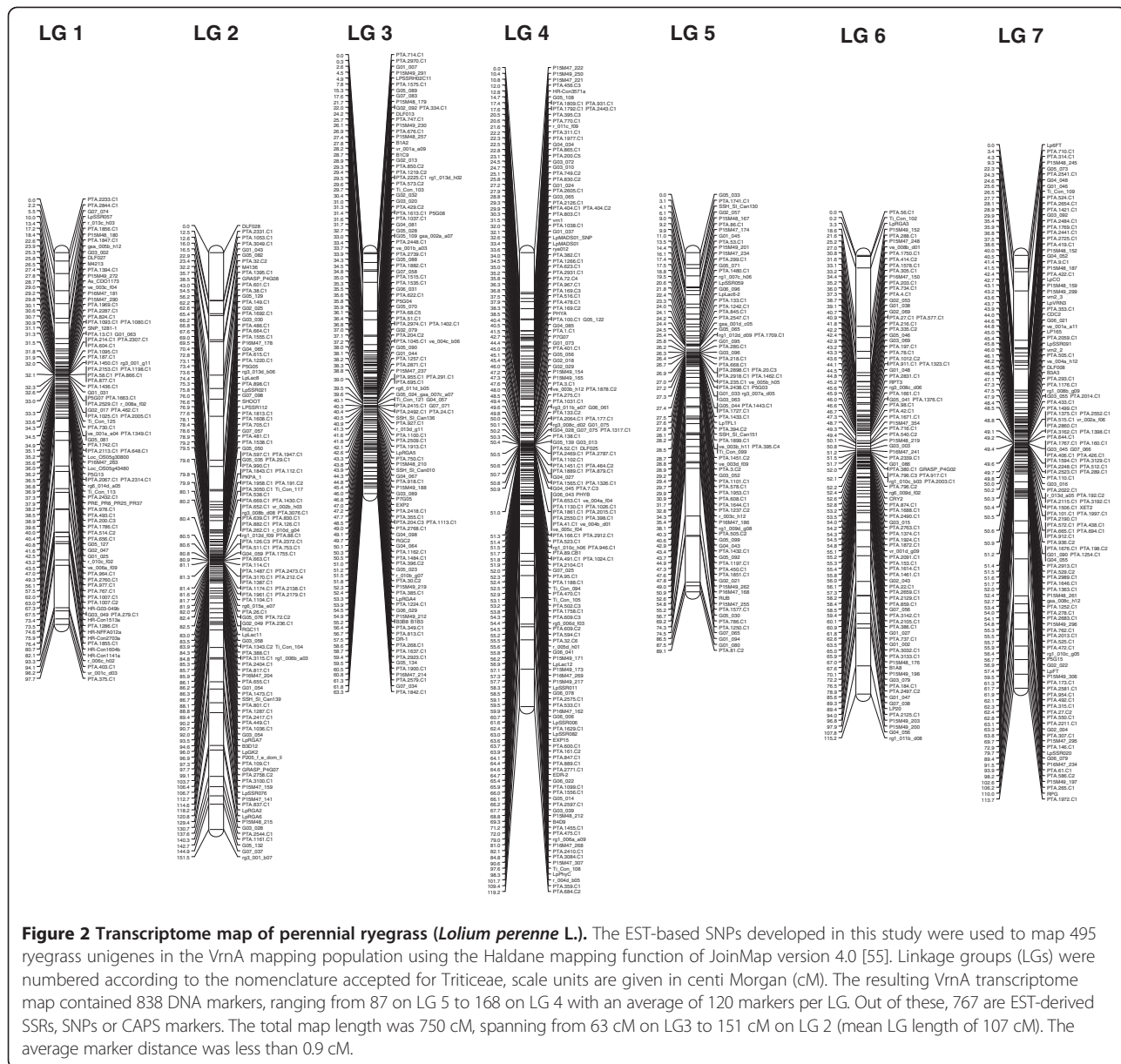
The two duplicated parental genotypes of the VrnA mapping population revealed highly consistent calls. For successfully genotyped SNPs, the frequency of missing values (MVs) was below 0.3% within the mapping population.

### Genetic linkage map

The mapping data of the VrnA map described in Jonavičienė et al. [17] and the 509 unigene SNPs were combined and grouped based on independence LOD scores. Markers were assigned to LGs at a LOD ratio threshold of 4.0 with the exception of LG1 and LG3, for which a LOD ratio threshold of 12 was necessary to separate the two LGs from each other. Fourteen SNPs failed to group with existing markers and were, therefore, excluded from mapping. Thus, a total of 495 SNP loci associated with transcribed genes (64% of the SNPs selected for GoldenGate genotyping) were located on the genetic linkage map (Additional file 2). The resulting VrnA map contained 838 DNA markers, ranging from 87 on LG 5 to 168 on LG 4 with an average of 120 markers per LG, of which a total of 767 are gene-derived SSRs, SNPs or CAPS markers (Figure 2). Markers were clustered around centromeric regions (Figure 2, Additional file 3: Figure S2). In order to estimate the accuracy of marker positions, 6 unigenes (PTA.1007.C, PTA.126.C1, PTA.404.C2 PTA.169.C3 PTA.609.C3 PTA.796.C3) were mapped based on more than one SNP. All SNPs derived



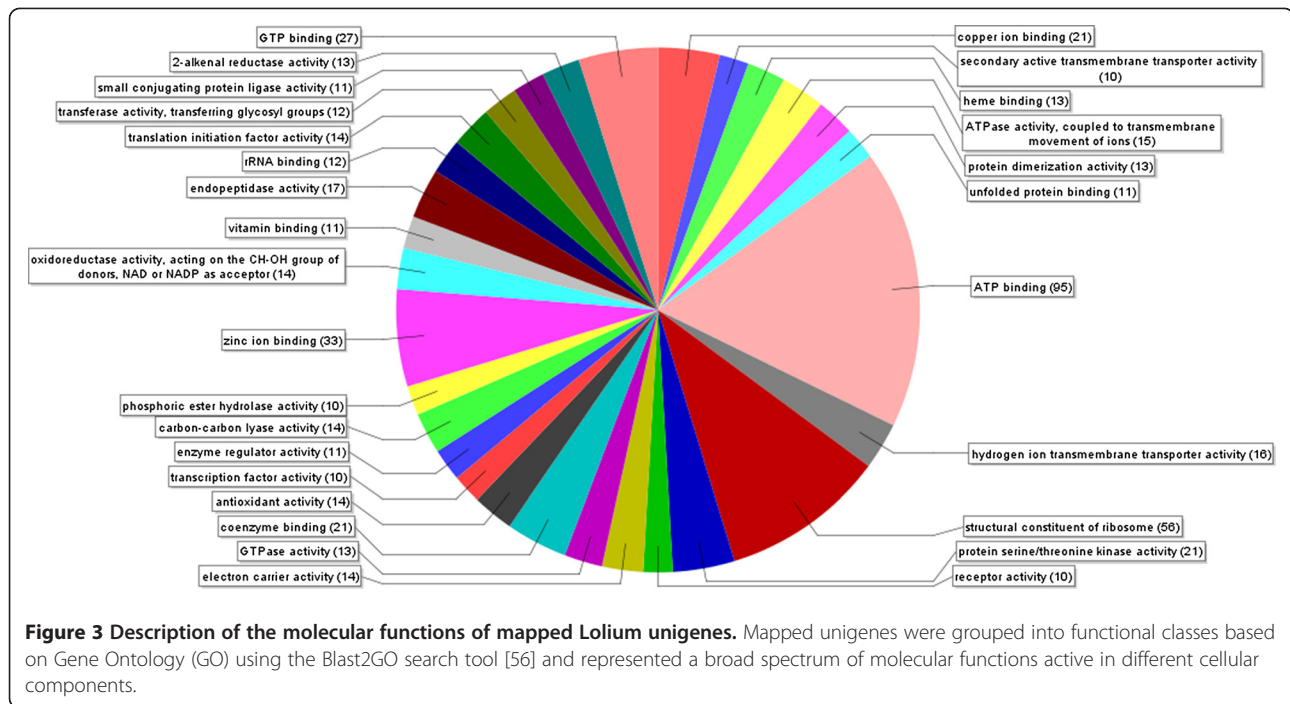
**Figure 1** Examples of SNP graphs observed in *Lolium oligo* pool assay (LOPA1) GoldenGate genotyping. SNP graphs are illustrated using the Software Illumina® GenomeStudio, version 2009.2. The normalized R (y-axis) is the normalized sum of intensities of the two dyes (Cy3 and Cy5), the normalized Theta (x-axis) is the deviation of Cy3 and Cy5 fluorescence from pure Cy3 and pure Cy5 signal (0 and 1). A normalized Theta value close to 0 and 1 is homozygous for SNP variant 1 and 2, respectively, a heterozygous sample is in between. The red, blue and purple ovals have the diameter of two standard deviations computed from the dispersal of the red, blue and purple dots, respectively. The numbers of plants in each cluster are indicated below the x-axis. **(A)** The 192 samples genotyped for SNP marker PTA.1021.C1 revealed fluorescence signal intensities close to 0, indicating assay failure. **(B)** Although the clustering algorithm at SNP PTA.1.C3 distinguished the three clusters at a GenTrain score of 0.40, such a genotyping pattern was considered inaccurate and this SNP was discarded from further analysis. **(C)** This illustration shows the SNP graph of monomorphic P9G02. **(D)** and **(E)** illustrate dominant SNPs being homozygous in one and heterozygous in the other mapping parent. For genetic linkage mapping, the markers PTA.109.C1 and PTA.291.C1 followed the segregation type  $nxnpx$  and  $lmxll$ , respectively [53]. Dots corresponding to the parents of the *VrnA* mapping population (which are represented in duplicates) are highlighted in yellow. Graph **(F)** shows a classical example of a SNP marker being heterozygous in both parents following the segregation pattern  $hxhkh$ .



**Figure 2** Transcriptome map of perennial ryegrass (*Lolium perenne* L.). The EST-based SNPs developed in this study were used to map 495 ryegrass unigenes in the VrnA mapping population using the Haldane mapping function of JoinMap version 4.0 [55]. Linkage groups (LGs) were numbered according to the nomenclature accepted for Triticeae, scale units are given in centiMorgan (cM). The resulting VrnA transcriptome map contained 838 DNA markers, ranging from 87 on LG 5 to 168 on LG 4 with an average of 120 markers per LG. Out of these, 767 are EST-derived SSRs, SNPs or CAPS markers. The total map length was 750 cM, spanning from 63 cM on LG3 to 151 cM on LG 2 (mean LG length of 107 cM). The average marker distance was less than 0.9 cM.

from the same unigene mapped with a distance of 1.5 cM, four of them within less than 0.2 cM. Similarly, two CAPS and a SNP marker derived from the *LpVrn1* gene mapped within less than 1.9 cM, whereas a CAPS and a SNP marker for *LpCO* mapped within the same cM. Another set of 18 SNPs were derived from unigenes previously mapped by EST-SSRs [54], allowing to compare performance and accuracy of SSR and SNP markers for genetic linkage mapping. Of the 18 comparisons, 11 (61%) mapped within 0.5 cM and only three that were located at the telomeric ends of the LGs, differed more than 3 cM. The slightly higher discrepancy of SNP and SSR map positions was an effect of the higher MV rate observed during SSR genotyping (data not shown).

Of the 732 non-redundant expressed genes mapped in VrnA, 654 (89%) revealed significant ( $E < e-10$ ) sequence similarities in a BLASTX search against the non-redundant (nr) protein database of GenBank, out of which 600 (82%) corresponded to genes with known molecular functions active in different cell components (Figure 3, Additional file 4: Figure S3, Additional file 5: Figure S4, Additional file 6: Figure S5). Unigenes were grouped in functional classes representing binding and catalytic activities (42% and 36%, respectively), structural molecule activities (8%), transport activities (7%), molecular transducer and transcription activities (2% each), enzyme regulatory activities (1%), as well as genes involved in nutrient uptake and transport (<1%).



The total map length was 750 cM, ranging from 63 cM on LG3 to 151 cM on LG 2 (mean LG length of 107 cM) with an average marker distance of less than 0.9 cM (Figure 2).

### Intra- and interspecific cross amplification

In addition to the *VrnA* mapping population including parental and grandparental genotypes, eight parental plants of four different perennial ryegrass mapping populations, one parent of the p150/112 intraspecific ILGI reference population [4] and the two parental genotypes of the Italian ryegrass (*Lolium multiflorum* Lam.) mapping population Xtg-ART [57] were used for genotyping. This allowed an estimation of the transferability of these SNPs to other genetic backgrounds. Of the 592 successfully genotyped SNPs, 275 (47%) detected reliable polymorphisms in at least one of the four additional perennial ryegrass mapping populations (between 201 and 250 for each population, Table 1), 48 of them (8%) were segregating in all populations. A total of 131 SNP markers (17%) detected polymorphisms segregating in Xtg-ART (Table 1). Interestingly, marker PTA.1032.C1 failed GoldenGate genotyping for perennial, but produced clear calls for the two Italian ryegrass plants. Markers PTA.32.CB2, PTA.43.C1, PTA.103.C1, PTA.271.C2, PTA.1535.C1, PTA.1613.C1, PTA.2333.C1, PTA.2371.C1 and r\_005b\_a08 were monomorphic in perennial ryegrass with a distinct genotype in Italian ryegrass. PTA.240.C2 and PTA.1044.C1 were monomorphic in perennial ryegrass but segregated in the Italian ryegrass mapping population Xtg-ART.

### Discussion

In recent years, technological advances in methods for high-throughput detection and genotyping of SNP markers have initiated a novel era in using molecular markers for genome analysis and breeding applications [58]. But still, the use of SNP markers for large-scale genome studies in allogamous forage grass species such as perennial ryegrass is still in its infancy. This is due to the low number of publicly available SNPs and the challenge of efficient SNP discovery and genotyping in a highly heterozygous genome containing a high proportion of repetitive elements and paralogous sequences. Here, we present both; an efficient SNP discovery pipeline based on 454 GS FLX transcriptome sequencing, and an Illumina GoldenGate assay to genotype, validate, and map the identified SNPs in the two way pseudo-testcross population *VrnA*.

### Genic SNP discovery in complex genomes

Transcriptome resequencing strategies and subsequent *in silico* SNP discovery have emerged as an efficient strategy for large-scale SNP discovery [29,37,58-63]. However, time and cost benefits are counterbalanced by a higher false discovery rate compared to *in vitro* approaches [64,65]. Incorrectly detected SNPs are primarily due to paralogous gene sequences interfering with the assembly of short NGS reads. In the present study, this was resolved by using a ryegrass unigene set with an average length of 889 bp as a reference for the assembly of the shorter 454 GS FLX transcriptome reads. The power of such an approach to

**Table 1 Intra- and interspecific cross amplification rates of SNPs on the *Lolium oligo* pool assay (LOPA1)**

| Species                        | Mapping population | Average call rate of all SNPs* | Number (percentage) of SNPs generating signals | Number (percentage) of polymorphic SNPs referred to SNPs generating signals | Number (percentage) of polymorphic SNPs referred to 768 SNPs on LOPA1 |
|--------------------------------|--------------------|--------------------------------|--|---|---|
| <i>Lolium perenne</i> L.       | VrnA               | 0.914 a                        | 692 (90%)                                      | 509 (74%)   | 509 (66%)   |
|                                | DLF1               | 0.846 b                        | 567 (74%)                                      | 241 (43%)   | 241 (31%)   |
|                                | DLF2               | 0.858 b                        | 605 (79%)                                      | 235 (39%)   | 235 (31%)   |
|                                | DLF3               | 0.840 b                        | 598 (78%)                                      | 201 (34%)   | 201 (26%)   |
|                                | DLF4               | 0.850 b                        | 601 (78%)                                      | 250 (42%)   | 250 (33%)   |
|                                | ILGI               | 0.882 c                        | 665 (87%)                                      | †192 (29%)  | 192 (25%)†  |
| <i>Lolium multiflorum</i> Lam. | Xtg-ART            | 0.818 d                        | 557 (73%)                                      | 131 (24%)   | 131 (17%)   |

\*Mapping populations with different letters vary significantly at  $P < = 0.05$  in their average call rates.

†Estimation from one parent only.

SNP performance of LOPA1 in VrnA was compared to different *Lolium multiflorum* Lam. (Xtg-ART) and *Lolium perenne* L. (DLF1 to 4) mapping populations, each represented with the two parental genotypes. For each mapping background, the average call rate of all SNPs is given and varies significantly at  $P < = 0.05$  if indicated with different letters. SNPs with a clear fluorescent signal detected in both mapping parents are considered as "SNPs generating signals". The numbers and percentages of polymorphic SNPs refer to SNP markers being heterozygous either in one or both parents, indicating a segregation pattern of *I*m*x*ll, *n*n*x*n*p* or *h*k*x*h*k* [53], respectively, in the corresponding mapping population. For the ILGI reference population represented with one parental genotype only, the number of SNPs being heterozygous is given. The percentages of polymorphic SNPs are referring to the number of SNPs generating signals and the total number of SNPs on LOPA1.

separate paralogous sequence variation has recently been shown in salmonids, whose genome contains a high degree of paralogous sequences due to a recent whole genome duplication event [66]. Moreover, a highly inbred ryegrass genotype was included for transcriptome sequencing as a means to identify paralogous genes and sequences from highly conserved domains of gene families in the alignment. As the inbred genotype was self-pollinated for six generations, the overall degree of heterozygosity is less than 1.5%. Genes that showed polymorphisms in reads from the inbred genotype indicated the presence of similar, non-allelic sequences and were therefore discarded for SNP discovery, thereby providing a reliable tool not only to reduce false positives in SNP discovery but also to facilitate the identification of genotype clusters during SNP genotyping.

Sequencing errors may represent an additional source of false positive SNPs. Even though error rates of NGS platforms are low (usually less than 1%) [67], a combination of Sanger sequencing (used for the establishment of the unigene set) and NGS (for transcriptome deep sequencing) was applied. Error rates of such combined sequencing approaches are even lower and thus an insignificant source of false-positive SNPs [68]. As a result, the present study revealed a false discovery rate (i.e., monomorphic SNP rate) of less than 12%, even lower than the initial estimation of 14%. The proportion of successfully called to finally mapped SNPs of 72% is comparable or slightly higher to validation rates between 57% and 77% observed in other species such as *Brachypodium distachyon* [69] or rye (*Secale cereale* L.) [63]. In conclusion, sequencing depth and a proper handling of paralogous sequences go hand in hand and are key factors for successful *in silico* SNP discovery approaches based on

RNA-seq. In future, large-scale NGS achieving longer read lengths and higher throughput in combination with improved assembly algorithms will provide opportunities for similar *in silico* SNP discovery approaches in less characterized species.

#### ***Lolium oligo* pool assays (LOPA1) design for ryegrass SNP genotyping**

Highly multiplexed Illumina SNP arrays are efficient tools to enhance mapping of expressed genes, thereby improving the resolution and usefulness of a genetic linkage map [42,48,69-73]. The use of a community OPA containing validated and well-performing SNPs as available for barley [42] is straightforward. However, the high calling rate (the rate of successfully genotyped SNPs) is often compromised by a lower conversion rate (the rate of polymorphic SNPs), as these SNPs were not a priori screened for polymorphisms within a particular mapping population. This was observed in barley, where approximately 51% of SNPs in the BOPA1 were polymorphic in a barley doubled haploid (DH) population [41]. Similarly, high calling (90%) but limited conversion rates (39 to 53%) were obtained when *de novo* OPA design was based on validated SNPs selected from public databases [48]. The percentage of polymorphic SNPs was even lower in *Pinus* and *Picea* species and ranged between 12 to 19% [65], which might be an effect of the very large and complex genomes [74], as well as limited sequence resources established for these species.

In contrast, much higher rates of polymorphic SNPs can be achieved by transcriptome resequencing of parental genotypes in the target mapping population, allowing the design of customized OPAs containing SNPs that are segregating in the mapping pedigree. While this was very

efficient to generate informative SNPs for linkage mapping, it might compromise the transferability of these SNPs to different genetic backgrounds. Given the high impact of additional polymorphisms in the flanking sequence of the target SNP on genotyping performance [75], intra- and interspecific SNP amplification rates in ryegrass might per se be lower when compared to inbreeding species due to increased nucleotide diversity present in outbreeding species. The detected 15,433 SNPs in 1,778 unigenes (this is an average of nine SNPs per unigene, one SNP every 102 bp) reflected the high nucleotide diversity present in a set of only four haplotypes. Nevertheless, the percentage of SNPs generating clear fluorescent signals (73 to 87%) was high in other Italian and perennial ryegrass backgrounds. Estimated rates of polymorphic SNPs ranging up to 33% indicate that LOPA1 can be applied to different genetic backgrounds. However, a more detailed study based on larger collections of various ryegrass genotypes will be required to confirm the significance of the reported SNP markers for broad-scale applications in ryegrasses. With the aim to further improve our *in silico* SNP discovery pipeline, the 76 SNPs failing GoldenGate genotyping were further examined and mapped back to genomic DNA. Interestingly, over 90% of these 76 SNPs had exon-intron boundaries within 20 bp flanking the target SNP (data not shown). This highlights an important drawback when developing SNPs from transcriptome sequencing data and indicates that BLASTN analysis to the rice genome sequence was inefficient to identify introns in ESTs for about 10% of the unigenes. A reference genome sequence will prove very useful to exactly locate intron-exon junctions for future large-scale SNP discovery studies.

#### Implications of the transcriptome map for ryegrass genetics and genomics

The ryegrass transcriptome map displays the genetic location of 732 expressed genes putatively underlying specific biochemical or physiological functions that control variation for agronomically important traits. The VrNA population has already proven to be valuable for mapping and cloning of major genes associated with meristem identity and the control of floral transition such as *LpVrm1*, *LpCO*, and *LpVrm3* [76,77]. For the same traits, the present transcriptome map contains additional candidate genes such as the TERMINAL FLOWER1-like gene (*LpTFL1*) that is a well characterised repressor of flowering and a controller of axillary meristem identity in ryegrass [78], and a homologue of the *Triticum monococcum* L. gene *TmVIL3*, that is up-regulated by vernalization [79]. The Arabidopsis homologue of *VIL3* is known to mediate chromatin modifications for stable repression of the *FLOWERING LOCUS C (FLC)*. Interestingly, the ryegrass homologue of *TmVIL3* (ve\_003c\_f04) mapped close to the centromere

on LG1, syntenic to the map position of *TmVIL3* in *T. monococcum*.

Another key trait that relates to vernalization response is fructan content, and the accumulation of fructans during cold acclimation. Fructans are known to play a key role in crop plants in response to abiotic stress in general, including drought, cold and freezing tolerance in particular [80]. In the present study, previously characterised, as well as novel genes involved in fructan biosynthesis were mapped, providing the opportunity to study fructan related metabolic processes involved in abiotic stress tolerance of grasses. This might be of particular interest since the VrNA grandparents – originating from different geographical latitudes – are not only significantly contrasting for their vernalization requirement, but also for the ability to accumulate fructans during cold acclimation, as well as in the response to drought treatment (unpublished data). Thus, given the high degree of segregation for traits such as abiotic stress tolerance and fructan accumulation in the VrNA population, it does represent a unique tool to unravel the gene regulatory networks of these traits.

Similarly, the current map contains genes involved in resistance to various biotic agents. Apart from the previously published NBS-LRR homologues [14,15], the map locates elements from disease resistance signal transduction pathways (Pto kinase interactor 1, p\_001c\_b08 corresponding to G02\_079) that were shown to be up-regulated after *Xanthomonas translucens* pv. *graminis* (Xtg) infection causing bacterial wilt [81]. Another gene showed high sequence similarity to members of the family of germin-like proteins (GLP; r\_010d\_c02) that are known to be involved in broad-spectrum basal defence against various pathogens and are also induced upon abiotic stress [82].

Other research groups can take advantage of this resource by using the unigene sequence information to develop simple 'Blind Mapping' HRM assays [77] to map a well distributed subset of the markers in their favourite mapping populations. This can then aid the transfer of information between different populations and species. The transcriptome map also serves as a source of candidate genes involved in various biological processes and molecular functions for association mapping. With an average marker distance of less than 0.9 cM, the presented VrNA map represents a good starting point for the establishment of BAC contigs for any genomic region of interest and will, in combination with the in-house BAC library established from one VrNA parental genotype [83], provide a very efficient toolbox for map-based cloning and gene isolation. However, it is worth noting that markers were not evenly distributed along the LGs, but clustered around the centromeres. Clustering of genes towards genetic centromeres due to low recombination frequencies is well known and has been



described in barley [84,85] and *Brachypodium* [69]. As a consequence, some markers at the centromeres could not be separated by 184 mapping individuals and co-segregated within recombination blocks. Thus, effects of MVs in mapping data became more apparent and single MVs resulted in slight changes of map positions, thereby explaining mapping discrepancies of two markers derived from the same unigene. We conclude that the current linkage map comes close to saturation of markers, at least in centromeric regions, and rather more mapping individuals than more markers would further improve map resolution. However, besides the general tendency that recombination frequency is reduced at genetic centromeres, it can vary dramatically along the chromosome [69]. *In silico* mapping of the unigene sequences to the ryegrass genome sequence, when available, will help resolve to what extent recombination frequencies vary along the chromosomes in greater detail, and will be valuable for ordering and orientation of scaffolds into pseudomolecules during the assembly of a ryegrass reference genome.

The availability of fully sequenced model grass genomes such as rice, *Brachypodium*, maize, and sorghum (*Sorghum bicolor* L. Moench) enables efficient exploitation of grass genome sequence resources for genetic and breeding applications in ryegrasses. Once established, syntenic relationships allow transferring map and marker information from related species across conserved genome regions [86]. Early comparative studies between the Pooideae tribes Triticeae and Poeae relied on restriction fragment length polymorphism (RFLP) markers mapped across different species and found that the genetic maps of perennial ryegrass and the Triticeae cereals are highly conserved in terms of orthology and colinearity [87,88]. However, these results were obtained from low-resolution genetic maps containing a limited number of anchor RFLP markers that allowed the detection of large rearrangements only, thereby missing a substantial part of the existing micro-synteny. Map and sequence-based markers presented here provide the opportunity to update and re-define synteny between ryegrass and the fully sequenced model grass genomes at a higher level of resolution to address micro-colinearity structure.

#### **Future prospective of high throughput SNP discovery and genotyping**

The advancements in sequencing and genotyping technology were a prerequisite for the work described here, and further improvements in throughput of NGS instruments can be expected. Combined with decreasing costs, it is worth considering genotyping by sequencing (GBS) approaches, thus by-passing the necessity for array-based genotyping [89]. In this case, we move straight to genotyping by means of sequencing all individuals of a mapping

or association panel. GBS strategies will prove extremely powerful for genome-wide association studies and for plant breeders moving towards implementing genomic selection in their breeding programmes [90].

However, whole genome resequencing may not be necessary when working within bi-parental mapping populations, where – depending on the population size – a finite amount of recombination and genome reshuffling is present. Thus, only SNP numbers adequate to cover the recombination blocks in the population are required. In this case, it may be sufficient to sequence a well distributed portion of the genome in all individuals [29]. A cost-effective approach of genotyping by sequencing on a small portion of the genome has recently been described and demonstrated in both maize and barley mapping populations [91]. The method described the use of a simple bar-coding strategy that allowed a high-level of multiplexing (up to 96-plex) and enabled mapping of approximately 200,000 and 25,000 sequence tags in maize and barley, respectively. With the increasing throughput of NGS, the authors envisage multiplexing up to 384 samples per lane, and thus pushing genotyping to under \$20 per sample. Although a reference genome is not necessarily required for this approach, it does allow for the use of genotype imputation methods when coverage is low.

Armed with these new powerful genotyping tools we can begin to reconsider how we construct mapping populations in order to improve power and precision. It will now be possible to densely genotype much larger populations for both bi-parental and association mapping studies, with the need for quality phenotyping remaining the sole bottleneck.

#### **Conclusions**

This study demonstrates the efficiency of using next generation transcriptome sequencing to discover gene-associated SNPs in species where no reference genome sequence has been established yet. In addition, we describe a workflow on how to successfully use the Illumina GoldenGate technology in outbreeding species characterized by highly heterozygous, large and complex genomes. We have also demonstrated the transferability of these SNPs to other perennial and Italian ryegrass mapping populations. The resulting map locates candidate genes for agronomically important traits and – at the given map resolution – represents a promising starting point for QTL fine mapping, LD-based association mapping, and map-based cloning via BAC clone isolation and sequencing. Moreover, the present EST map provides new anchor points for detailed studies of comparative grass genomics that will prove useful for future ordering and orientation of scaffolds into pseudomolecules during the assembly of a ryegrass reference genome.

## Methods

### Mapping population

The VrnA two-way pseudo-testcross mapping population consisting of 184 F2 perennial ryegrass genotypes [2] was used to map the EST-derived SNPs. These plants were complemented with eight parental genotypes of four different perennial ryegrass mapping populations, one parent of the p150/112 intraspecific ILGI reference population [4], and two Italian ryegrass plants which have been used to establish the Xtg-ART population characterized for bacterial wilt and crown rust resistance [57,92]. Genomic DNA was isolated from young leaves following a phenol/chloroform extraction protocol with minor modifications described in Jensen et al. [2].

### RNA isolation

Total RNA from both parents of the VrnA population (NV#20 F1-30 and NV#20 F1-39, respectively) as well as the inbred genotype p226/179/2 was isolated using Tri<sup>®</sup> Reagent (Sigma-Aldrich, St. Louis, MO, USA) according to the manufacturer's instructions. Isolation of mRNA and synthesis of cDNA was performed according to Milano et al. [38].

### SNP discovery

The unigene set was generated according to Asp et al. [52] using the PHRED, PHRAP and CROSS\_MATCH software packages [93-95]. For the final assembly, the PHRAP minmatch threshold was 75, all other parameters were set to default. The Roche FLX 454 technology was used to generate reads using barcoded libraries [96] from NV#20 F1-30, NV#20 F1-39 and the inbred genotype p226/179/2. The alignment of the 454 reads to the unigene set was based on the Mosaik sequence assembler (<http://bioinformatics.bc.edu/marthlab/Mosaik/>). A hash size of 15 was used with a mismatch threshold set to a maximum of 4% mismatches. Large-scale SNP detection in the assembled contigs was performed using GigaBayes V0.4.1 [97] with a minimum of four total reads at each SNP position and a minimum read coverage of two for each SNP variant. Minimum base quality was 10, the probability threshold of each SNP at least 0.5.

### SNP validation

Prior to GoldenGate assay design, a subset of detected SNPs were validated by HRM or direct sequencing of PCR products amplified from the parental genotype being heterozygous for the target SNP. For HRM analysis, a total of twelve mapping individuals along with the parental genotypes were used for short amplicon melting as described by Studer et al. [77]. Primers used for short amplicon melting were designed to flank the target SNP with an amplicon product size of 40 to 60 bp.

Sequencing of PCR fragments was performed at Eurofins MWG Operon, Ebersberg, Germany.

### Development of the *Lolium oligo pool assay* (LOPA1)

LOPA1 used in this study consisted of 786 SNPs selected according to the following criteria: (i) heterozygosity of the target SNP in one or both parental genotypes of VrnA, (ii) absence of additional polymorphisms adjacent to the target SNP, (iii) the detected SNPs were located within a distance of 50 bp to sequence ends or intron/exon splice junctions (iv), absence of polymorphism in sequence reads of the highly inbred reference genotype p226/179/2 within a contig and (v) Illumina assay design score > 0.6 as determined by the Illumina Technical Service. The final set of 768 SNPs addressed 760 ryegrass unigenes, out of which eight were covered with two SNPs.

### SNP genotyping

The parental genotypes of the VrnA mapping population were genotyped in duplicate. Genotyping was performed according to the manufacturer's protocol on 96-well format Sentrix arrays [98] using the BeadArray technology in combination with an allele-specific extension, adapter ligation and amplification assay protocol. Arrays were imaged using a BeadArray Reader Scanner. Genotyping data generated by the Software Illumina<sup>®</sup> GenomeStudio, version 2009.2 were manually inspected and corrected for misclassification of genotypes.

### Linkage analysis and map construction

The genetic linkage map of the VrnA population illustrated in Jonavičienė et al. [17] was complemented with 509 gene-associated SNPs. Markers were assigned to LGs using independence LOD scores for group formation. Map construction was based on regression mapping at LOD and recombination threshold value of 1.00 and 0.40, respectively, using the software package JoinMap 4.0 [55]. Map distances were calculated using the Haldane mapping function implemented in JoinMap 4.0.

The annotation of mapped unigenes, including a thorough description of their molecular functions, biological processes and cell compartments involved, was determined based on Gene Ontology (GO) using the Blast2GO search tool [56].

### Heat map construction

The marker density from the ryegrass transcriptome map was visualized by counting the number of markers in a window size of 3 cM shifted in 0.3 cM steps along a LG using a manual python script. Color scale was adapted to the minimum (dark blue = 0 marker/3 cM) and maximum (red = 17 to 52 marker/3 cM) window counts, adjusted for each LG separately.

## Additional files

**Additional file 1: Figure S1. SNP validation by high resolution melting (HRM) curve analysis (A) and direct sequencing of PCR fragments (B).** (A) shows the normalized melting curves of a target SNP for twelve mapping individuals along with the parental genotypes that were used for short amplicon melting as described by Studer et al. [77]. The melting curves given in grey represent individuals being homozygous for the target SNP, while red melting curves indicated heterozygous individuals. The sequencing trace file given in (B) illustrates the results from direct sequencing of PCR products amplified from the parental genotype being heterozygous for the target SNP. Sequencing of PCR fragments was performed at Eurofins MWG Operon, Ebersberg, Germany.

**Additional file 2: Detailed description of SNP markers.** This table contains the unigene names and GenBank accession numbers along with detailed mapping information (the linkage group and map position) and the SNP polymorphism used for GoldenGate genotyping.

**Additional file 3: Figure S2. Heat map of DNA markers on the perennial ryegrass transcriptome map.** Marker density on each linkage group (LG) was visualized as heat maps by counting the number of markers in a window of 3 centi Morgan (cM) size shifted in 0.3 cM steps along a LG using an in-house python script. Color scale was adapted to the minimum (dark blue = 0 marker/2 cM) and maximum (red = 17 to 52 marker/3 cM) window counts, adjusted for each LG separately.

**Additional file 4: Figure S3. Summary of unigene annotation.** The 732 non redundant *Lolium* unigenes were subjected to a BLASTN search against the non-redundant (nr) nucleotide database of Genbank, mapped and functionally annotated based on Gene Ontology (GO) using the Blast2GO search tool [56].

**Additional file 5: Figure S4. Description of biological processes affected by mapped *Lolium* unigenes.** Biological processes were determined based on Gene Ontology (GO) using the Blast2GO search tool [56]. The number of mapped unigenes involved in a specific process is given in parenthesis.

**Additional file 6: Figure S5. Description of cellular components involved in molecular functions of mapped *Lolium* unigenes.** Mapped unigenes were allocated to cellular components based on Gene Ontology (GO) using the Blast2GO search tool [56]. The number of unigenes for each cellular component is given in parenthesis.

## Competing interests

The authors declare that they have no competing interests.

## Acknowledgements

The authors would like to acknowledge Illumina Technical Service, Jonas Grauholm and Thomas Thykjaer from AROS Applied Biotechnology A/S, as well as Stephan Hentrup from the Department of Molecular Biology and Genetics at Aarhus University for excellent technical support. This study was funded by The Danish Council for Independent Research, Technology and Production Sciences (project 274-08-0300), and partly supported by the Danish Directorate for Food, Fisheries and Agri Business (project FRØMARK, 3412-05-01313).

## Author details

<sup>1</sup>Department of Molecular Biology and Genetics, Faculty of Science and Technology, Research Centre Flakkebjerg, Aarhus University, Forsøgsvej 1, 4200 Slagelse, Denmark. <sup>2</sup>Department of Molecular Biology and Genetics, Faculty of Science and Technology, Research Centre Foulum, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark. <sup>3</sup>Institute of Bioinformatics and Systems Biology, Helmholtz Center Munich, German Research Center for Environmental Health, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany. <sup>4</sup>Department of Agronomy, Iowa State University, 1204 Agronomy Hall, 50011 Ames, IA, USA.

## Authors' contributions

BS, TL and TA conceived the study. TA extracted RNA, CB and FP coordinated the sequencing and performed SNP discovery. TA and BS designed LOPA1, MSI and BS validated selected SNPs prior to GoldenGate genotyping. BS

coordinated the GoldenGate assay, extracted the mapping data and performed the linkage mapping. BS drafted the manuscript, which was improved by TA, SB, MP and TL. All authors read and approved the final manuscript.

Received: 15 December 2011 Accepted: 18 April 2012

Published: 18 April 2012

## References

1. Jones ES, Dupal MP, Dumsday JL, Hughes LJ, Forster JW: **An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne* L.).** *Theor Appl Genet* 2002, **105**:577–584.
2. Jensen LB, Andersen JR, Frei U, Xing Y, Taylor C, Holm PB, Lübberstedt T: **QTL mapping of vernalization response in perennial ryegrass (*Lolium perenne* L.) reveals co-location with an orthologue of wheat *VRN1*.** *Theor Appl Genet* 2005, **110**:527–536.
3. Muylle H, Baert J, Van Bockstaele E, Pertijs J, Roldán-Ruiz I: **Four QTLs determine crown rust (*Puccinia coronata* f. sp. *lolii*) resistance in a perennial ryegrass (*Lolium perenne*) population.** *Heredity* 2005, **95**:348–357.
4. Bert PF, Charmet G, Sourdil P, Hayward MD, Balfourier F: **A high-density molecular map for ryegrass (*Lolium perenne*) using AFLP markers.** *Theor Appl Genet* 1999, **99**:445–452.
5. Armstead IP, Turner LB, King IP, Cairns AJ, Humphreys MO: **Comparison and integration of genetic maps generated from F<sub>2</sub> and BC<sub>1</sub>-type mapping populations in perennial ryegrass.** *Plant Breed* 2002, **121**:501–507.
6. Barre P, Mi F, Balfourier F, Ghesquière M: **QTLs for morphogenetic traits and sensitivity to rusts in *Lolium perenne*.** In *Proceedings of the Second International Symposium on Molecular Breeding of Forage Crops; Lorne and Hamilton, Victoria, Australia, November 19–24, 2000*. Edited by Spangenberg G. 2000: 60.
7. van Loo EN, Dolstra O, Humphreys MO, Wolters L, Luessink W, de Riek W, Bark N: **Lower nitrogen losses through marker assisted selection for nitrogen use efficiency and feeding value (NIMGRASS).** *Vort Pflanz* 2003, **59**:270–279.
8. Anhalt U, Heslop-Harrison JP, Byrne S, Guillard A, Barth S: **Segregation distortion in *Lolium*: evidence for genetic effects.** *Theor Appl Genet* 2008, **117**:297–306.
9. Faville MJ, Vecchies AC, Schreiber M, Drayton MC, Hughes LJ, Jones ES, Guthridge KM, Smith KF, Sawbridge T, Spangenberg GC, et al: **Functionally associated molecular genetic marker map construction in perennial ryegrass (*Lolium perenne* L.).** *Theor Appl Genet* 2004, **110**:12–32.
10. Lauvergeat V, Barre P, Bonnet M, Ghesquière M: **Sixty simple sequence repeat markers for use in the *Festuca-Lolium* complex of grasses.** *Mol Ecol Notes* 2005, **5**:401–405.
11. Kopecky D, Bartos J, Lukaszewski A, Baird J, Cernoch V, Kölliker R, Rognlí OA, Blois H, Caig V, Lübberstedt T, et al: **Development and mapping of DArT markers within the *Festuca - Lolium* complex.** *BMC Genomics* 2009, **10**:473.
12. Studer B, Asp T, Frei U, Hentrup S, Meally H, Guillard A, Barth S, Muylle H, Roldán-Ruiz I, Barre P, et al: **Expressed sequence tag-derived microsatellite markers of perennial ryegrass (*Lolium perenne* L.).** *Mol Breed* 2008, **21**:533–548.
13. Jensen LB, Muylle H, Arens P, Andersen CH, Holm PB, Ghesquière M, Julier B, Lübberstedt T, Nielsen KK, Riek JD, et al: **Development and mapping of a public reference set of SSR markers in *Lolium perenne* L.** *Mol Ecol Notes* 2005, **5**:951–957.
14. Schejbel B, Jensen LB, Xing Y, Lübberstedt T: **QTL analysis of crown rust resistance in perennial ryegrass under conditions of natural and artificial infection.** *Plant Breed* 2007, **126**:347–352.
15. Schejbel B, Jensen LB, Asp T, Xing Y, Lübberstedt T: **Mapping of QTL for resistance to powdery mildew and resistance gene analogues in perennial ryegrass (*Lolium perenne* L.).** *Plant Breed* 2008, **127**:368–375.
16. Studer B, Jensen LB, Hentrup S, Brazauskas G, Kölliker R, Lübberstedt T: **Genetic characterisation of seed yield and fertility traits in perennial ryegrass (*Lolium perenne* L.).** *Theor Appl Genet* 2008, **117**:781–791.
17. Jonavičienė K, Studer B, Asp T, Jensen LB, Paplauskienė V, Lazauskas S, Brazauskas G: **Identification of genes involved in a 6-days water deprivation response in timothy (*Phleum pratense* L.) and mapping of orthologous loci in perennial ryegrass (*Lolium perenne* L.).** *Biol Plantarum* 2012, **56**:473–483.
18. Rafalski A: **Applications of single nucleotide polymorphisms in crop genetics.** *Curr Opin Plant Biol* 2002, **5**:94–100.

19. Ponting RC, Drayton MC, Cogan NOI, Dobrowolski MP, Spangenberg GC, Smith KF, Forster JW: **SNP discovery, validation, haplotype structure and linkage disequilibrium in full-length herbage nutritive quality genes of perennial ryegrass (*Lolium perenne* L.).** *Mol Gen Genomics* 2007, **278**:585–597.
20. Gupta PK, Rustgi S, Mir RR: **Array-based high-throughput DNA markers for crop improvement.** *Heredity* 2008, **101**:5–18.
21. Hamblin MT, Buckler ES, Jannink J-L: **Population genetics of genomics-based crop improvement methods.** *Trends Genet* 2011, **27**:98–106.
22. Ganai MW, Altmann T, Röder MS: **SNP identification in crop plants.** *Curr Opin Plant Biol* 2009, **12**:211–217.
23. Cogan NOI, Ponting RC, Vecchies AC, Drayton MC, George J, Dracatos PM, Dobrowolski MP, Sawbridge TI, Smith KF, Spangenberg GC, Forster JW: **Gene-associated single nucleotide polymorphism discovery in perennial ryegrass (*Lolium perenne* L.).** *Mol Gen Genomics* 2006, **276**:101–112.
24. Buetow KH, Edmonson MN, Cassidy AB: **Reliable identification of large numbers of candidate SNPs from public EST data.** *Nat Genet* 1999, **21**:323–325.
25. Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M: **Mining SNPs From EST Databases.** *Genome Res* 1999, **9**:167–174.
26. Somers DJ, Kirkpatrick R, Moniwa M, Walsh A: **Mining single-nucleotide polymorphisms from hexaploid wheat ESTs.** *Genome* 2003, **46**:431–437.
27. Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Graner A: **Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.).** *Mol Gen Genomics* 2003, **270**:24–33.
28. Morozova O, Marra MA: **Applications of next-generation sequencing technologies in functional genomics.** *Genomics* 2008, **92**:255–264.
29. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML: **Genome-wide genetic marker discovery and genotyping using next-generation sequencing.** *Nat Rev Genet* 2011, **12**:499–510.
30. Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J: **A first-generation haplotype map of maize.** *Science* 2009, **326**:1115–1117.
31. Deschamps S, Rota ML, Ratashak JP, Biddle P, Thureen D, Farmer A, Luck S, Beatty M, Nagasawa N, Michael L: **Rapid genome-wide single nucleotide polymorphism discovery in soybean and rice via deep resequencing of reduced representation libraries with the Illumina genome analyzer.** *The Plant Genome* 2010, **3**:53–68.
32. Myles S, Chia J-M, Hurwitz B, Simon C, Zhong GY, Buckler E, Ware D: **Rapid genomic characterization of the genus *Vitis*.** *PLoS ONE* 2010, **5**:e8219.
33. Hyten DL, Song Q, Fickus EW, Quigley CV, Lim J-S, Choi I-Y, Hwang E-Y, Pastor-Corrales M, Cregan PB: **High-throughput SNP discovery and assay development in common bean.** *BMC Genomics* 2010, **11**:475.
34. Hyten DL, Cannon SB, Song Q, Weeks N, Fickus EW, Shoemaker RC, Specht JE, Farmer AD, May GD, Cregan PB: **High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence.** *BMC Genomics* 2010, **11**:38.
35. Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS: **SNP discovery via 454 transcriptome sequencing.** *Plant J* 2007, **51**:910–918.
36. Trick M, Long Y, Meng J, Bancroft I: **Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing.** *Plant Biotechnol J* 2009, **7**:334–346.
37. Barbazuk WB, Schnable PS: **SNP discovery by transcriptome pyrosequencing.** *Methods Mol Biol* 2011, **729**:225–246.
38. Milano I, Babbucci M, Panitz F, Ogden R, Nielsen RO, Taylor MI, Helyar SJ, Carvalho GR, Espiñeira M, Atanassova M, et al: **Novel tools for conservation genomics: Comparing two high-throughput approaches for SNP discovery in the transcriptome of the European hake.** *PLoS ONE* 2011, **6**:e28008.
39. Fan JB, Chee MS, Gunderson KL: **Highly parallel genomic assays.** *Nat Rev Genet* 2006, **7**:632–644.
40. Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, et al: **Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties.** *Proc Natl Acad Sci USA* 2006, **103**:18656–18661.
41. Sato K, Takeda K: **An application of high-throughput SNP genotyping for barley genome mapping and characterization of recombinant chromosome substitution lines.** *Theor Appl Genet* 2009, **119**:613–619.
42. Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, et al: **Development and implementation of high-throughput SNP genotyping in barley.** *BMC Genomics* 2009, **10**:582.
43. Hyten D, Song Q, Choi I-Y, Yoon M-S, Specht J, Matukumalli L, Nelson R, Shoemaker R, Young N, Cregan P: **High-throughput genotyping with the GoldenGate assay in the complex genome of soybean.** *Theor Appl Genet* 2008, **116**:945–952.
44. Blanc G, Wolfe KH: **Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes.** *Plant Cell* 2004, **16**:1667–1678.
45. Choi I-Y, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon M-S, et al: **A soybean transcript map: Gene distribution, haplotype and single-nucleotide polymorphism analysis.** *Genetics* 2007, **176**:685–696.
46. Sandve SR, Rudi H, Dørum G, Berg PR, Rognli OA: **High-throughput genotyping of unknown genomic terrain in complex plant genomes: lessons from a case study.** *Mol Breed* 2010, **26**:711–718.
47. Lai J, Li R, Xu X, Jin W, Xu M, Zhao H, Xiang Z, Song W, Ying K, Zhang M, et al: **Genome-wide patterns of genetic variation among elite maize inbred lines.** *Nat Genet* 2010, **42**:1027–1030.
48. Yan J, Yang X, Shah T, Sánchez-Villeda H, Li J, Warburton M, Zhou Y, Crouch J, Xu Y: **High-throughput SNP genotyping with the GoldenGate assay in maize.** *Mol Breed* 2010, **25**:441–451.
49. Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J: **Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers.** *PLoS ONE* 2009, **4**:e8451.
50. Akhunov E, Nicolet C, Dvorak J: **Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay.** *Theor Appl Genet* 2009, **119**:507–517.
51. Durstewitz G, Polley A, Plieske J, Luerssen H, Graner EM, Wieseke R, Ganai MW: **SNP discovery by amplicon sequencing and multiplex SNP genotyping in the allopolyploid species *Brassica napus*.** *Genome* 2010, **53**:948–956.
52. Asp T, Frei UK, Didion T, Nielsen KK, Lübberstedt T: **Frequency, type, and distribution of EST-SSRs from three genotypes of *Lolium perenne*, and their conservation across orthologous sequences of *Festuca arundinacea*, *Brachypodium distachyon*, and *Oryza sativa*.** *BMC Plant Biol* 2007, **7**:36.
53. Maliepaard C, Jansen J, Van Ooijen JW: **Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications.** *Genet Res* 1997, **70**:237–250.
54. Studer B, Kölliker R, Muylle H, Asp T, Frei U, Roldán-Ruiz I, Barre P, Tomaszewski C, Meally H, Barth S, et al: **EST-derived SSR markers used as anchor loci for the construction of a consensus linkage map in ryegrass (*Lolium spp.*).** *BMC Plant Biol* 2010, **10**:177.
55. Van Ooijen JW: **JoinMap<sup>®</sup> 4, Software for the calculation of genetic linkage maps in experimental populations.** Kyazma BV, Wageningen, Netherlands 2006.
56. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**:3674–3676.
57. Studer B, Boller B, Herrmann D, Bauer E, Posselt UK, Widmer F, Kölliker R: **Genetic mapping reveals a single major QTL for bacterial wilt resistance in Italian ryegrass (*Lolium multiflorum* Lam.).** *Theor Appl Genet* 2006, **113**:661–671.
58. Varshney RK, Nayak SN, May GD, Jackson SA: **Next-generation sequencing technologies and their implications for crop genetics and breeding.** *Trends Biotechnol* 2009, **27**:522–530.
59. Bancroft I, Morgan C, Fraser F, Higgins J, Wells R, Clissold L, Baker D, Long Y, Meng JL, Wang XW, et al: **Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing.** *Nat Biotechnol* 2011, **29**:762–766.
60. Edwards D, Batley J: **Plant genome sequencing: applications for crop improvement.** *Plant Biotechnol J* 2010, **8**:2–9.
61. Imelfort M, Duran C, Batley J, Edwards D: **Discovering genetic polymorphisms in next-generation sequencing data.** *Plant Biotechnol J* 2009, **7**:312–317.
62. Jackson SA, Iwata A, Lee SH, Schmutz J, Shoemaker R: **Sequencing crop genomes: approaches and applications.** *New Phytol* 2011, **191**:915–925.
63. Haseneyer G, Schmutz T, Seidel M, Zhou R, Mascher M, Schön C-C, Taudien S, Scholz U, Stein N, Mayer KFX, Bauer E: **From RNA-seq to large-scale genotyping - genomics resources for rye (*Secale cereale* L.).** *BMC Plant Biol* 2011, **11**:131.

64. Lepoittevin C, Frigerio J-M, Garnier-Géré P, Salin F, Cervera M-T, Vornam B, Harvengt L, Plomion C: *In vitro vs in silico* detected SNPs for the development of a genotyping array: What can we learn from a non-model species? *PLoS ONE* 2010, **5**:e11034.
65. Chancerel E, Lepoittevin C, Le Provost G, Lin Y-C, Jaramillo-Correa JP, Eckert AJ, Wegrzyn JL, Zelenika D, Boland A, Frigerio J-M, et al: Development and implementation of a highly-multiplexed SNP array for genetic mapping in maritime pine and comparative mapping with loblolly pine. *BMC Genomics* 2011, **12**:368.
66. Everett MV, Grau ED, Seeb JE: Short reads and nonmodel species: exploring the complexities of next-generation sequence assembly and SNP discovery in the absence of a reference genome. *Mol Ecol Resour* 2011, **11**:93–108.
67. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM: Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* 2007, **8**:R143.
68. You FM, Huo N, Deal KR, Gu YQ, Luo M-C, McGuire PE, Dvorak J, Anderson OD: Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. *BMC Genomics* 2011, **12**:59.
69. Huo N, Garvin DF, You FM, McMahon S, Luo M, Cheng, Gu YQ, Lazo GR, Vogel JP: Comparison of a high-density genetic linkage map to genome features in the model grass *Brachypodium distachyon*. *Theor Appl Genet* 2011, **123**:455–464.
70. Deulvot C, Charrel H, Marty A, Jacquin F, Donnadiou C, Lejeune-Henaut I, Burstin J, Aubert G: Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. *BMC Genomics* 2010, **11**:468.
71. Pavy N, Pelgas B, Beauseigle S, Blais S, Gagnon F, Gosselin I, Lamothe M, Isabel N, Bousquet J: Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC Genomics* 2008, **9**:21.
72. Hyten DL, Choi I-Y, Song Q, Specht JE, Carter TE, Shoemaker RC, Hwang E-Y, Matukumalli LK, Cregan PB: A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. *Crop Sci* 2010, **50**:960–968.
73. Anithakumari AM, Tang J, van Eck HJ, Visser RGF, Leunissen JAM, Vosman B, van der Linden C: A pipeline for high throughput detection and mapping of SNPs from EST databases. *Mol Breed* 2010, **26**:65–75.
74. Ahuja MR: Recent advances in molecular genetics of forest trees. *Euphytica* 2001, **121**:173–195.
75. Grattapaglia D, Silva OB, Kirst M, de Lima BM, Faria DA, Pappas GJ: High-throughput SNP genotyping in the highly heterozygous genome of Eucalyptus: assay success, polymorphism and transferability across species. *BMC Plant Biol* 2011, **11**:65.
76. Andersen JR, Jensen LB, Asp T, Lübberstedt T: Vernalization response in perennial ryegrass (*Lolium perenne* L.) involves orthologues of diploid wheat (*Triticum monococcum*) *VRN1* and Rice (*Oryza sativa*) *Hd1*. *Plant Mol Biol* 2006, **60**:481–494.
77. Studer B, Jensen LB, Fiil A, Asp T: "Blind" mapping of genic DNA sequence polymorphisms in *Lolium perenne* L. by high resolution melting curve analysis. *Mol Breed* 2009, **24**:191–199.
78. Jensen CS, Salchert K, Nielsen KK: A terminal flower1-like gene from perennial ryegrass involved in floral transition and axillary meristem identity. *Plant Physiol* 2001, **125**:1517–1528.
79. Fu D, Dunbar M, Dubcovsky J: Wheat *VIN3*-like PHD finger genes are up-regulated by vernalization. *Mol Gen Genomics* 2007, **277**:301–313.
80. Livingston DT, Hinch DK, Heyer AG: Fructan and its relationship to abiotic stress tolerance in plants. *Cell Mol Life Sci* 2009, **66**:2007–2023.
81. Wichmann F, Asp T, Widmer F, Kölliker R: Transcriptional responses of Italian ryegrass during interaction with *Xanthomonas translucens* pv. *graminis* reveal novel candidate genes for bacterial wilt resistance. *Theor Appl Genet* 2011, **122**:567–579.
82. Manosalva PM, Davidson RM, Liu B, Zhu X, Hulbert SH, Leung H, Leach JE: A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. *Plant Physiol* 2009, **149**:286–296.
83. Farrar K, Asp T, Lübberstedt T, Xu ML, Thomas AM, Christiansen C, Humphreys MO, Donnison IS: Construction of two *Lolium perenne* BAC libraries and identification of BACs containing candidate genes for disease resistance and forage quality. *Mol Breed* 2007, **19**:15–23.
84. Stein N, Prasad M, Scholz U, Thiel T, Zhang H, Wolf M, Kota R, Varshney R, Perovic D, Grosse I, Graner A: A 1,000-loci transcript map of the barley genome: new anchoring points for integrative grass genomics. *Theor Appl Genet* 2007, **114**:823–839.
85. Mayer KFX, Martis M, Hedley PE, Šimková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, et al: Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell* 2011, **23**:1249–1263.
86. Bennetzen JL, Freeling M: Grasses as a single genetic system: genome composition, collinearity and compatibility. *Trends Genet* 1993, **9**:259–261.
87. Sim S, Chang T, Curley J, Warnke SE, Barker RE, Jung G: Chromosomal rearrangements differentiating the ryegrass genome from the *Triticeae*, oat, and rice genomes using common heterologous RFLP probes. *Theor Appl Genet* 2005, **110**:1011–1019.
88. Jones ES, Mahoney NL, Hayward MD, Armstead IP, Jones JG, Humphreys MO, King IP, Kishida T, Yamada T, Balfourier F, et al: An enhanced molecular marker based genetic map of perennial ryegrass (*Lolium perenne*) reveals comparative relationships with other Poaceae genomes. *Genome* 2002, **45**:282–295.
89. Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, et al: High-throughput genotyping by whole-genome resequencing. *Genome Res* 2009, **19**:1068–1076.
90. Meuwissen T, Goddard M: Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics* 2010, **185**:623–631.
91. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE: A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 2011, **6**:e19379.
92. Studer B, Boller B, Bauer E, Posselt U, Widmer F, Kölliker R: Consistent detection of QTLs for crown rust resistance in Italian ryegrass (*Lolium multiflorum* Lam.) across environments and phenotyping methods. *Theor Appl Genet* 2007, **115**:9–17.
93. Ewing B, Hillier L, Wendl MC, Green P: Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998, **8**:175–185.
94. Ewing B, Green P: Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998, **8**:186–194.
95. Gordon D, Abajian C, Green P: Consed: A graphical tool for sequence finishing. *Genome Res* 1998, **8**:195–202.
96. Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willerslev E: The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* 2007, **2**:e197.
97. Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, Zakeri H, Stitzel NO, Hillier L, Kwok P-Y, Gish WR: A general approach to single-nucleotide polymorphism discovery. *Nat Genet* 1999, **23**:452–456.
98. Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, et al: Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol* 2003, **68**:69–78.

doi:10.1186/1471-2164-13-140

Cite this article as: Studer et al.: A transcriptome map of perennial ryegrass (*Lolium perenne* L.). *BMC Genomics* 2012 **13**:140.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

