RNA-seq analysis reveals genetic response and tolerance mechanisms to ozone exposure in soybean

Adam Whaley1, Jaime Sheridan1, Sajedeh Safari1, Amy Burton2, Kent Burkey2 and Jessica Schlueter1*

Abstract

Background: Oxidative stress caused by ground level ozone is a contributor to yield loss in a number of important crop plants. Soybean (Glycine max) is considered to be ozone sensitive, and current research into its response to oxidative stress is limited. To better understand the genetic response in soybean to oxidative stress, an RNA-seq analysis of two soybean cultivars was performed comparing an ozone intolerant cultivar (Mandarin-Ottawa) and an ozone resistant cultivar (Fiskeby III) following exposure to ozone.

Results: Analysis of the transcriptome data revealed cultivar-specific expression level differences of genes previously implicated in oxidative stress responses, indicating unique cultivar-specific responses. Both Fiskeby III and Mandarin (Ottawa) exhibit an increased expression of oxidative response genes as well as glutathiones, phenylpropanoids, and phenylalanine ammonia-lyases. Mandarin (Ottawa) exhibited more general stress response genes whereas Fiskeby III had heightened expression of metabolic process genes. An examination of the timing of gene responses over the course of ozone exposure identified significantly more differentially expressed genes across all time points in Mandarin (Ottawa) than in Fiskeby III. The timing of expression was also considered to identify genes that may be indicative of a delayed response to ozone stress in Fiskeby III. We found that Mandarin (Ottawa) exhibits an higher level of expression in early time points for oxidative and general stress response genes while Fiskeby III seems to maintain expression of defense and stress response genes. Of particular interest was the expression of wax and cutin biosynthetic genes that we found to be expressed in Mandarin (Ottawa) in all sampled time points, whereas the expression of this pathway is only in the first time point for Fiskeby III.

Conclusions: We were able to identify differentially expressed genes that correspond to each of the known or expected categories of genes previously implicated in other species for ozone stress. Our study shows evidence that at least part of the observed ozone tolerance of Fiskeby III may be due to its thicker, denser leaves providing passive resistance thereby limiting the degree of ozone exposure. The observed diminished genetic response is then likely a consequence of this reduced exposure.

Background

Ozone pollution is often perceived as an urban issue, but the problem is, in reality, regional and includes many agricultural areas. This problem occurs in rural areas because nitrogen oxides and hydrocarbons produced by electric power generation or urban sources such as automobiles move as plumes into rural areas. Those precursors then react with oxygen in the air and sunlight to form ozone [1]. The impact of ozone on human health receives significant public attention, but many plants are also sensitive to this toxic air pollutant [2–4].

As an economically important crop, soybean suffers from several abiotic stresses that severely affect its production and quality in most countries [5,6]. However, studies to identify genetic materials that are tolerant to abiotic stress, and specifically ozone stress, have been limited in soybean. Fortunately, recent studies have identified a small group of soybean genotypes from a breeding program in Fiskeby, Sweden, that are tolerant to several of the environmental stresses to which they have been challenged [7]. These
closely related soybean genotypes are all tolerant or partially tolerant to drought, iron deficiency chlorosis, toxic soil aluminum, salt, cold, and atmospheric ozone pollution. Burkey and Carter (2009) found that these Fiskeby varieties were significantly more tolerant to ozone stress than many other soybean varieties [7].

Ozone has been shown to reduce growth and yield in many crops, and soybean is known to be an ozone sensitive species that experiences stippling and necrosis of the leaves under elevated ozone conditions [8–12]. Leaf tissue damage is the combined result of activated oxygen species and an ethylene stress response that is interpreted as a pathogen invasion leading to cell death as a means to limit the spread of the perceived pathogen [13]. Impacts on a specific farm in a specific year can be difficult to quantify due to variations in the frequency and intensity of ozone pollution during the growing season and from year to year. Modeling studies that combine measurement of ambient ozone concentration with empirically derived yield response curves suggest that current soybean yields are reduced up to 10 % by ozone pollution [14, 15]. Research at the SoyFACE project in Illinois, where soybeans are exposed in the field to elevated ozone levels, predicts an additional 20 % yield reduction by the year 2050 if ambient ozone levels rise [16]. These yield losses could be exacerbated due to the current urbanization trend.

Ozone exposure directly affects the leaves by entering through the stomata where it is converted to reactive oxygen species including hydrogen peroxide, superoxide anions and hydroxyl radicals [17]. These reactive oxygen molecules are responsible for inducing the oxidative stress response. Upon exposure to ozone, plants may employ several response mechanisms some of which are conserved among species. Following exposure, ozone has an affect on photosynthesis resulting in an increase in carbon dioxide concentration that in turn reduces stomatal conductance [18]. Once ozone has entered a leaf, much of the defense response is thought to focus on either preventing the formation of reactive oxygen signaling molecules or reducing the concentration of plant-derived reactive oxygen molecules once they form. Superoxide dismutase, peroxidase and glutathione in conjunction of ascorbic acid work to remove these reactive oxygen molecules from the cell [19–21]. Expression studies in tobacco showed an increase in beta-1,3-glucanase, catalase and chitinase [22]. Physiologically, as the stomates are closing there is a decrease in mRNA levels of both subunits of Rubisco in potato plants exposed to ozone [23]. Runneckles and Chevone described increased respiration, decreased photosynthesis, peroxidation of the membrane lipids, reduced transpiration, closing of the stomates and increased senescence [24]. Ethylene and salicylic acid, signaling molecules for oxidative stress, are negatively regulated by jasmonic acid and act as both a source of resistance and susceptibility to ozone by regulating plant cell death [25].

In soybean, much like pathogen response, ozone elicits a hypersensitive resistance response that leads to a heightened production of flavonoids; in turn, this increase in flavonoids is thought to be a key factor in toxicity in the leaves that appears in the form of stippling and necrosis [10]. Maccarrone et al. showed that in soybean, lipoxygenase (LOX) transcription is increased in response to ozone exposure [26]. More recent work by Galant et al. supported the importance of the redox pathways in response to ozone exposure [27]. Using a proteomics approach, Galant et al. also showed that oxidative stress alters the thiol oxidative state of proteins and leads to increased expression of peroxidase, methionine, and glutathione [27]. Ozone stress QTLs have been identified in rice, but limited QTL studies to date have addressed this stress in soybeans [28]. Although a study undertaken by Burton et al. has led to the discovery of several putative ozone QTLs, the genetic mechanisms for ozone tolerance in soybean are largely unknown.

Within soybean, an important clue to understanding and dissecting these stress responses is found in the unique germplasm, geography, and geology of Fiskeby, Sweden, the site of a soybean breeding program that produced varieties from the 1940’s through ~1975. Fiskeby is near Stockholm, Sweden and is far north of other soybean growing areas of the world leading to the production of early maturing varieties adapted to the short growing season. The stress-resistant nature of the varieties released by the Swedish program, Fiskeby III (USDA Plant Introduction (PI) 438471) being an example, was not discovered until recently when genetic studies included screening the ancestral base of North American varieties (Fiskeby types being part of the ancestral base for Canada) for various traits [7]. Resistance to ozone may have been a by-product of Holmberg’s selection of varieties adapted to the cool climate of Sweden. It has been postulated that thicker leaves associated with cold tolerant Fiskeby genotypes could contribute to ozone tolerance by mechanisms yet to be identified, however testing of this hypothesis is ongoing.

In order to identify the genetic components of ozone resistance, we conducted an RNA-seq analysis of tolerant and sensitive soybean cultivars under both control (low ozone) and stressed (high ozone) conditions. We focused on early response to ozone stress as we hypothesize that this point is when gene expression in Fiskeby III supports tolerance mechanisms.

**Results and discussion**

Although Holmberg bred several lines in his program that exhibited abiotic stress tolerance, Fiskeby III was chosen as the representative for this project because it exhibited the greatest ozone tolerance of the Fiskeby...
germplasm tested in a follow-up open-top chamber field trial after the initial greenhouse screening [7]. Similarly, Fiskeby III was chosen as one of the parental lines in a recombinant inbred population derived from a cross with Mandarin (Ottawa), an abiotic stress sensitive line being used by Burkey and Carter to investigate several abiotic stresses including ozone tolerance. This will facilitate the eventual comparison and integration of results from both classical genetic and molecular genomic investigative methods.

RNA-seq libraries were sequenced on the Illumina HiSeq2000 and generated 40 million to 200 million reads per lane (Fig. 1). Following quality control, Tophat mapped between 84–95 % of reads back to Glycine max Wm82.a2.v1 reference genome [29,30]. For each sample, only a small number of reads, ranging from ~3,000 to 72,000 reads failed to meet QC standards in the Tophat mapping process. Read quality figures for read duplication as well as dispersion were also created to assess read quality. (see Additional file 1: Figure S1, S2).

Analysis of RNA-seq results to determine ozone response in ozone-sensitive soybean

Using cuffdiff [31], differential analysis of the sequencing results of the Mandarin (Ottawa) at low and high ozone concentrations results yielded 135 to 275 differentially expressed genes per time point, or less than 1 % of the total number of annotated genes in the reference genome (Table 1). In total, 535 unique genes showed differential expression at one or more time points. Genes with annotated functions or processes matching phenylpropanoids, and lignin biosynthesis were found in all four time points, while genes corresponding to glutathione, lipoxygenase, and phenylalanine ammonia-lyases were found in one or more time points but not in all four time points. Of the entire set of genes differentially expressed in all four time points 267 genes had a consistent pattern of up or down regulation across all four time points. Of these genes, 230 showed higher expression in the high ozone condition, and 15 correspond with GO functions related to oxidative or general stress response in soybean: Glyma.09G156700, Glyma.11G078400 and Glyma.17G036200 are putative oxidative stress response genes that code for peroxidase genes. Glyma.05G231900 is a gene involved in phenylpropanoid biosynthesis, but also match GO terms for systemic acquired resistance, a pathway in plants that “warns” surrounding plant tissues after a localized pathogen exposure, making its expression here particularly interesting. Glyma.10G162400 is a transcription-regulating gene that matches GO terms.
implicating it in auxin, cadmium, and osmotic stress response. The remaining 10 genes, Glyma.02G142500, Glyma.04G003200, Glyma.04G088500, Glyma.04G248500, Glyma.05G109600, Glyma.07G018000, Glyma.10G001800, Glyma.10G179400, Glyma.14G141000, Glyma.14G216200, all match GO terms related to salt stress, and other terms such as jasmonic acid response, response to cold, cell wall biogenesis, and oxidation-reduction (Fig. 2; see Additional file 1: Table S1). Because of the prevalence of these pathways in the results, it is likely that many of the genes in salt and cold stress response also play a yet uncharacterized role in oxidative stress response.

In addition to the enrichment of expected oxidative stress responses in ozone exposed samples, an analysis of individual time points using gene ontology (GO) term enrichment showed a shifting enrichment in transcriptional regulation over the time course. During time points 1–3, a higher proportion of differentially expressed genes came from larger FPKM values in the high ozone exposure samples (165:100, 92:43, 214:1, respectively), and the high exposure sample had a larger number of genes matching GO terms related to salt stress, and other terms such as jasmonic acid response, response to cold, cell wall biogenesis, and oxidation-reduction (Fig. 2; see Additional file 1: Table S1). Because of the prevalence of these pathways in the results, it is likely that many of the genes in salt and cold stress response also play a yet uncharacterized role in oxidative stress response.

Analysis of RNA-seq results to determine ozone response in ozone-tolerant soybean
Differential analysis of the sequencing results of Fiskeby III samples comparing low to high ozone exposure showed fewer differentially expressed genes than the Mandarin (Ottawa) samples; the number of differentially expressed genes ranged from 88 to 384 genes, a similar number of genes to the previous analysis of Mandarin (Ottawa) at low and high ozone. (Table 1). In contrast to the Mandarin (Ottawa) series, only lignin biosynthesis and lipoxygenase showed some degree of differential expression across all four time points. Phenylalanine ammonia-lyases showed expression in the first two time points only. Phenylpropanoid genes showed differential expression in all but the final time point, while glutathione had minimal differential expression with only one instance of differential expression, occurring at the final time point. 470 unique genes were found to have differential expression in one or more time point, and these genes did show a significant GO term abundance for oxidation reduction when analyzed through AgriGO, (Fig. 3). When the entire time series was examined as a whole, however, 1682 genes were found to have a consistent pattern of differential expression, by far the largest number of differentially expressed genes in any sample comparison. Of these 1682 genes, 863 had higher expression in the high ozone samples. Significant GO terms returned from AgriGO were mostly related to lipid, carbohydrate, and fatty acid metabolic processes. However, it did also show a significant number of GO terms related to negative regulation of molecular function (p = 3.7e–5). There were a number of GO terms related to oxidative stress, however the p-value for those terms was not statistically significant (p = 0.051). An expression heatmap of these genes can be found in Fig. 4 (see Additional file 1: Table S2).

Comparison of tolerant and sensitive genotypes
Comparing the high ozone exposure between both Mandarin (Ottawa) and Fiskeby III at individual time points resulted in 258 to 536 differentially expressed genes, a similar number to the comparisons of each
Lipoxygenase and lignin genes were found at all four time points, whereas phenylpropanoids and glutathione genes were found at time points 2–4. Phenylalanine ammonia-lyases were found at the first, third, and fourth time points. The 258 to 536 differentially expressed genes per time point correspond to 564 unique genes with differential expression at a minimum of one time point. An AgriGO analysis of these 564 genes identified a significant abundance of GO terms related to oxidation-reduction (p = 5.1e−6). When the time series was examined as a whole, only 64 genes were shown to have a consistent pattern of expression across all four time points. Of these 64 genes, 58 had higher expression in the ozone sensitive Mandarin (Ottawa), and only 6 had higher expression in Fiskeby. Of the 6 genes that had higher expression in Fiskeby, we were only able to identify GO terms for four. Glyma.14G164900 and Glyma.14G165000 matched identical GO terms for oxidation-reduction and response to light stimulus, and are likely tandemly duplicated genes. Glyma.11G197300 is a cytochrome P450 gene that matches a large number of biological process GO terms including defense response, response to water deprivation, and induced systemic resistance. Glyma.08G287500 matches GO terms for cell wall organization, anthocyanin accumulation, and regulation of hormone levels. 58 genes had higher expression in Mandarin (Ottawa), and 24 of these matched GO terms related to oxidative or general stress response. A summary of these genes and a brief description of their matching biological process GO terms can be found in Table 2. Of particular note in these results are the two transcription regulators, Glyma.13G203700 and Glyma.15G182600 which matched GO terms for regulation of hydrogen peroxide and systemic acquired resistance. The previously described cultivar at low and high exposure described above. Lipoxygenase and lignin genes were found at all four time points, whereas phenylpropanoids and glutathione genes were found at time points 2–4. Phenylalanine ammonia-lyases were found at the first, third, and fourth time points. The 258 to 536 differentially expressed genes per time point correspond to 564 unique genes with differential expression at a minimum of one time point. An AgriGO analysis of these 564 genes identified a significant abundance of GO terms related to oxidation-reduction (p = 5.1e−6). When the time series was examined as a whole, only 64 genes were shown to have a consistent pattern of expression across all four time points. Of these 64 genes, 58 had higher expression in the ozone sensitive Mandarin (Ottawa), and only 6 had higher expression in Fiskeby. Of the 6 genes that had higher expression in Fiskeby, we were only able to identify GO terms for four. Glyma.14G164900 and Glyma.14G165000 matched identical GO terms for oxidation-reduction and response to light stimulus, and are likely tandemly duplicated genes. Glyma.11G197300 is a cytochrome P450 gene that matches a large number of biological process GO terms including defense response, response to water deprivation, and induced systemic resistance. Glyma.08G287500 matches GO terms for cell wall organization, anthocyanin accumulation, and regulation of hormone levels. 58 genes had higher expression in Mandarin (Ottawa), and 24 of these matched GO terms related to oxidative or general stress response. A summary of these genes and a brief description of their matching biological process GO terms can be found in Table 2. Of particular note in these results are the two transcription regulators, Glyma.13G203700 and Glyma.15G182600 which matched GO terms for regulation of hydrogen peroxide and systemic acquired resistance. The previously described
SAR gene, Glyma.05G231900 was expressed in both cultivars and not deemed significantly different.

While these results highlight some of the differences in the oxidative stress responses between tolerant and sensitive genotypes, also of interest is the timing of responses as the samples transition from one time point to the next when comparing the two cultivars at high ozone exposure. One posited explanation for the ozone tolerance seen in the Fiskeby cultivar was a delay or limited ozone uptake through some yet uncharacterized mechanism. This made an examination of the timing of differential gene expression between cultivars of particular interest. We examined a number of patterns (Table 3) and identified a number of different genes that showed delayed or accelerated expression in each cultivar. The 39 genes that showed an accelerated expression in Mandarin (Ottawa) were not sufficient to generate significant GO terms from AgriGO, but could still be examined individually. Of these 39, there were 11 that could be implicated as part of oxidative or general stress response. These include
3 oxidation-reduction process genes, Glyma.07G225300, Glyma.10G203500, and Glyma.20G051700. There were also three genes that matched GO terms relating to transcriptional control, Glyma.08G194900, Glyma.19G035300, and Glyma.10G082500. This last gene, Glyma.10G082500, is of particular interest, as it matched with over 52 biological process GO terms, many of which are implicated in oxidative stress and other stress response pathways. The analysis of genes matching a pattern of accelerated expression in Fiskeby, 27 genes in total, was also insufficiently large to produce significant GO abundances from AgriGO. In addition to identifying fewer genes matching this pattern of expression, there were far fewer genes within the results related to oxidative stress. Only Glyma.03G162700, an ethylene signaling pathway gene, and Glyma.18G087000, a defense response gene, matched GO terms that could be attributed to stress response. The apparent lack of an accelerated response from the ozone tolerant plant is one of the more compelling results that lend evidence to the hypothesis that the physical characteristics of the Fiskeby leaf limit the severity of ozone exposure and therefore limit leaf damage.

Another pattern of expression examined were the genes that were expressed in both cultivars at the first time point at high ozone exposure but were turned off at a later time point in one cultivar while the other maintained expression under high ozone conditions. Both possible configurations of this pattern had a small number of genes associated with it: 65 genes in which expression was maintained in Fiskeby III, 39 genes in which expression was maintained in Mandarin (Ottawa) (Table 3). In the first scenario, where expression is maintained in Fiskeby III but reduced or eliminated in Mandarin (Ottawa), these genes were overwhelmingly related to defense or stress response. Of these genes, 10 matched GO terms related to systemic acquired resistance, 4 additional genes matched GO terms for oxidative stress response, 7 more were related to oxidation-reduction, and

![Heatmap of FPKM values of oxidative stress genes in Fiskeby low and high exposure across all time points. A heatmap of FPKM values for 9 genes identified by agrIGO as oxidative stress genes. Strength of expression (FPKM) is indicated by color, ranging from very low yellow to deep red at high FPKM values. The genes shown were found to be significantly differentially expressed by cufflinks (p < 0.05), but were not sufficient to pass an agrIGO analysis for statistically significant GO term enrichment (p = 0.051).](image-url)
| Glyma Identifier | Biological Process GO terms |
|------------------|----------------------------|
| Glyma.02G205800  | cell wall biogenesis; cell wall macromolecule metabolic process;  cellulose biosynthetic process;  hydrogen peroxide biosynthetic process;  plant-type cell wall biogenesis; |
| Glyma.03G145600  | defense response to nematode;  oxidation-reduction process;  response to oxidative stress |
| Glyma.03G222000  | response to abscisic acid stimulus;  response to cold;  response to heat;  response to high light intensity;  response to hydrogen peroxide;  response to oxidative stress;  response to salt stress;  response to water deprivation |
| Glyma.04G063800  | defense response to bacterium;  defense response to fungus;  plant-type cell wall biogenesis;  response to osmotic stress;  secondary cell wall biogenesis |
| Glyma.04G088500  | response to salt stress |
| Glyma.04G227700  | flavonol biosynthetic process;  lignin biosynthetic process;  phenylpropanoid metabolic process;  positive regulation of flavonoid biosynthetic process;  response to wounding |
| Glyma.06G065000  | cellulose biosynthetic process;  defense response to bacterium;  defense response to fungus;  glucuronoxylan metabolic process;  plant-type cell wall biogenesis;  positive regulation of abscisic acid biosynthetic process;  response to osmotic stress;  response to water deprivation;  secondary cell wall biogenesis;  xylan biosynthetic process |
| Glyma.06G158800  | hydrogen peroxide biosynthetic process;  proteolysis;  transition metal ion transport |
| Glyma.07G139000  | lignin biosynthetic process;  lignin catabolic process;  oxidation-reduction process;  phenylpropanoid metabolic process;  plant-type cell wall biogenesis |
| Glyma.07G157100  | defense response;  lignan biosynthetic process |
| Glyma.09G005600  | Golgi organization;  hyperosmotic response;  lipid metabolic process;  protein targeting to vacuole;  proteolysis;  response to salt stress;  response to temperature stimulus |
| Glyma.09G038500  | carbohydrate metabolic process;  cell wall macromolecule catabolic process;  glucuronoxylan metabolic process;  hydrogen peroxide biosynthetic process;  lignin biosynthetic process;  protein desumoylation;  vegetative to reproductive phase transition of meristem;  xylan biosynthetic process |
| Glyma.09G051100  | cell wall biogenesis;  defense response to bacterium;  defense response to fungus |
| Glyma.10G035500  | regulation of transcription, DNA-dependent |
| Glyma.11G053200  | glucuronoxylan metabolic process;  secondary cell wall biogenesis |
| Glyma.12G233700  | response to endoplasmic reticulum stress;  response to heat;  response to high light intensity;  response to hydrogen peroxide;  signal transduction |
| Glyma.13G094900  | plant-type cell wall organization;  response to chitin;  response to cold;  response to heat;  response to mechanical stimulus;  response to wounding |
| Glyma.13G203700  | plant-type cell wall modification;  regulation of transcription, DNA-dependent;  transmitting tissue development |
| Glyma.13G301900  | response to wounding |
| Glyma.15G110200  | Golgi organization;  hyperosmotic response;  lipid metabolic process;  protein targeting to vacuole;  response to cadmium ion;  response to salt stress;  response to temperature stimulus |
| Glyma.15G143600  | carbohydrate metabolic process;  cell wall macromolecule catabolic process;  glucuronoxylan metabolic process;  hydrogen peroxide biosynthetic process;  lignin biosynthetic process;  protein desumoylation;  vegetative to reproductive phase transition of meristem;  xylan biosynthetic process |
| Glyma.15G182600  | regulation of hydrogen peroxide metabolic process;  response to hypoxia;  systemic acquired resistance;  salicylic acid mediated signaling pathway |
| Glyma.17G072200  | cell wall biogenesis;  hydrogen peroxide biosynthetic process |
| Glyma.19G008200  | pollen tube growth;  regulation of defense response |

6 others related to jasmonic acid or ethylene response. In addition to there being fewer genes with expression maintained in Mandarin (Ottawa), the genes that did match this pattern did not show the same prevalence as the previous pattern towards stress or defense responses. Only one gene in this set, Glyma.01G153300 was related to oxidation-reduction, and no genes matched GO terms for oxidative stress response. One additional gene matched GO terms
Furthermore, ozone is more readily deposited due to the overall lower number of differential gene calls, which in turn causes there to be too few genes to identify the pathway through AgriGO or other GO term abundance tools. In the initial pattern analysis that revealed the cuticle wax pathway we observed 143 genes where we now only observe 43. With the exception of the Fiskeby low exposure to Fiskeby high ozone exposure time series comparison, all other comparisons saw a decrease in the number of differentially expressed genes of a similar magnitude. Additionally, the exclusion of cuticle wax genes from the pattern results could be also attributed to the new genome or annotations as well; only 66 genes in the new annotations match GO terms for wax biosynthesis. If we examine the individual gene annotations from the pattern script, there are several genes we can attribute to the wax biosynthesis pathway. These include two ketoacyl-CoA synthases (Glyma.10G179400 and Glyma.0U033000), which are involved in fatty acid elongation for cuticle wax formation [39]. There is also a wax ester synthase gene (Glyma.06G291700). Using the 66 genes matching GO terms for wax biosynthesis, we were able to query the data for those genes specifically and were able to identify three genes that showed differential expression. Of these genes, Glyma.11G185100, Glyma.12G183400 showed significant differential expression at time point 2, and had a fold change corresponding to higher expression in Fiskeby at that time point. However, by time point four the fold change had switched to correspond to higher expression in Mandarin (Ottawa), although it was not found to be significant (log2(fold change) = 8.925). The other gene, Glyma.13G317600, showed differential expression at time points one, two, and four. At the first two time points, the fold change showed higher expression in Fiskeby, though by time point four Glyma.13G317600 had higher expression in Mandarin (Ottawa). All three of these genes substantiate the previously observed pattern of expression, where initial expression favors Fiskeby or neither cultivar, and over the time points shifts to higher expression in Mandarin (Ottawa). Additionally, we also used a list of Arabidopsis cuticle wax biosynthesis genes from Le Provost et al., a study that linked cuticle wax and drought tolerance [40]. Soybean orthologs to the A. thaliana genes from the Le Provost study were identified using BLAST, and the cufflinks results for those genes, in addition to the 66 genes matching GO terms for wax biosynthesis, can be seen in Supplemental Table 3 (See Additional File 1). While only a limited number of these genes have statistically significant differential expression, there is a strong overall trend that matches that seen in the pattern 3 genes. Specifically, we see more genes with fold changes values corresponding to higher expression in Fiskeby in the first time point, which decreases to more genes showing higher expression in Mandarin (Ottawa) by the final time point.

| Number of genes matching pattern | Pattern Description                                                                 |
|---------------------------------|------------------------------------------------------------------------------------|
| 34                              | Starts with 0 expression in both cultivars, faster response in Mandarin (Ottawa)    |
| 27                              | Starts with 0 expression in both samples, faster response in Fiskeby               |
| 65                              | Starts with expression in both samples, expression lowers in Mandarin (Ottawa) while Fiskeby remains expressed |
| 39                              | Starts with expression in both samples, expression lowers in Fiskeby while Mandarin (Ottawa) remains expressed |

A description of each gene expression pattern examined as part of the expression timing analysis. The number of genes that matched a particular pattern is also included.

related to jasmonic acid response and several abiotic stress responses, another matched terms related to hydrogen peroxide response.

**Analysis of wax and cutin biosynthesis pathway**

One pathway of particular interest to our investigation was the cuticle wax biosynthesis pathway. During our initial probing of the data using Cufflinks 2.0.2 and the soybean reference genome (version 1.89), we identified an abundance of genes belonging to the cuticle wax biosynthesis pathway that matched expression pattern 3 (expression in both cultivars, expression tapers or ends in Fiskeby). This result was unexpected for two reasons: first, cuticle wax biosynthesis is not well documented as a part of ozone tolerance and has only been documented in a limited set of species; and secondly, if it is down regulated in the ozone tolerant cultivar, what can the presence of this pathway tell us about ozone tolerance. A literature review of ozone tolerance in other plant species yielded some promising information regarding its possible role in ozone tolerance and resistance in plant species. In plums, cranberries and spruce, ozone exposure results in a significant decrease in cuticle depth [33–35]. Furthermore, ozone is more readily deposits onto the leaf surface when the cuticular layer is reduced leading to greater foliar injury in response to ozone, and ozone has been previously implicated in the degradation of cuticular wax [36, 37]. Work by Zhao et al. showed that the application of an exogenous chitosan relieved some of the damage caused by ozone exposure in soybean [38].

When the read mapping was updated for the new genome version, using Cufflinks 2.2.1, the script for determining differentially expressed genes matching certain patterns no longer returned the same abundance of wax biosynthesis pathway genes that our previous analysis had. At this time, we hypothesize that this is likely due to the overall lower number of differential gene calls,
Since literature supports the idea of the cuticular wax pathway being an important component of ozone tolerance, then the critical question becomes why do we observe this pattern of expression that favors the ozone intolerant cultivar? To answer this question we reviewed the physical characteristics of the leaves of each cultivar. Table 4 shows genotypic differences in leaf physiology in response to ozone treatment in both Fiskeby III and Mandarin (Ottawa). Leaf foliar injury, leaf air space and the specific weight of leaves all confirm that Mandarin (Ottawa) is exhibiting greater ozone sensitivity than Fiskeby III. The leaf air space data shows that there is much more exposure of cell surface area in leaves from Mandarin (Ottawa), a potential basis for the greater ozone sensitivity. The specific leaf weight data shows that there is more cellular mass in Fiskeby III leaves as a result of the leaves being either thicker and/or having more densely packed cells.

Reviewing these characteristics reveals what we believe to be a critically important factor when attempting to explain the expression of cuticular wax pathways in the ozone intolerant Mandarin (Ottawa). What becomes immediately evident from the physical data is that the leaves of the ozone tolerant Fiskeby cultivar are more massive and the cell density of the leaf tissue is much higher than that in Mandarin (Ottawa). We believe that because of this the ability for ozone to penetrate the leaf tissue and inflict injury upon the leaf tissue is severely limited. Because Mandarin (Ottawa) lacks this specific benefit, we see expression of cuticle wax pathways in an attempt to bolster the physical barrier to ozone penetration of leaf tissues. This may also explain the differences in anti-oxidant expression between the two cultivars: if Fiskeby, by virtue of its thicker leaf structure, does not need to express cuticular wax biosynthesis in order to build a physical barrier, than perhaps it is able to devote those transcriptional resources towards anti-oxidant pathways. And because this response is already dealing with a less significant influx of active oxygen species compared to that in Mandarin (Ottawa), it is able to significantly limit the degree of leaf damage the plant suffers.

**Conclusion**

The analysis of the RNA-seq data generated from a time series exposure of ozone tolerant Fiskeby and ozone intolerant Mandarin (Ottawa) soybean varieties was remarkable not just for the genes identified during the study, but also for how small the set of expressed genes identified were. Only one comparison made was able to find differential expression in more than 1 % of soybean genes. This observation is likely due to combination of factors, the relatively short ozone exposure times means we are only observing initial gene expression responses and likely the start of response pathways. We also must consider that the low ozone controls means we can reliably say that pathways and genes unrelated to oxidative stress would not be differentially expressed when we compare the low and high exposure sample sets of each cultivar. When comparing the high exposure samples for each cultivar to each other, it is likely that the low number of differentially expressed genes is a function of how similar the two cultivars are, which in turn limits the number of genes which would be expressed differently between the two cultivars. Despite the small number of differentially expressed genes in each comparison set, we were able to identify differentially expressed genes that correspond to each of the known or expected categories of genes identified in other species. Despite the presence of the expected responses in each cultivar, the timing of expression of these expected genes and responses is distinct between each cultivar, and likely this is at least partly responsible for the difference in the

| Physiological Parameter | Genotype          | Treatment (ppb O₃) | Leaf 3 | Leaf 4 | Leaf 5 |
|------------------------|-------------------|--------------------|--------|--------|--------|
| Foliar Injury (% leaf area) | Fiskeby III | 27 | 0 | 0 | 0 |
| N = 4 | Mandarin (Ottawa) | 74 | 0 | 0 | 0 |
| Air Space (ml gFW⁻¹) | Fiskeby III | 27 | 0.49 ± 0.02 | n/a | 0.45 ± 0.02 |
| N = 4-6 | Mandarin (Ottawa) | 74 | 0.50 ± 0.02 | n/a | 0.43 ± 0.02 |
| Specific weight (gDW m⁻²) | Fiskeby III | 27 | n/a | 57 ± 2 | n/a |
| N = 6 | Mandarin (Ottawa) | 74 | n/a | 53 ± 1 | n/a |
| Mandarin (Ottawa) | 74 | n/a | 31 ± 1 | n/a |
outcome from ozone exposure. Another important finding, related to the small number of genes, is the apparent lack of response observed in Fiskeby in both the high exposure comparison to Mandarin (Ottawa) in both the number of differentially expressed genes and also in the gene expression pattern analysis. The fact that we observe only 3 genes in Fiskeby related to oxidative stress expressed first in Fiskeby before Mandarin (Ottawa), as well as the comparison of Fiskeby and Mandarin (Ottawa) at high ozone which only returned 6 genes which showed higher expression over the entire time series lead us to conclude that there is an overall diminished genetic response to ozone exposure in the ozone tolerant variety compared to the ozone intolerant variety. When combined with the evidence of cuticle wax expression in the ozone intolerant variety and the differences in the physical characteristics of each leaf, we believe that at least part of the observed ozone tolerance of Fiskeby is due to its thicker, denser leaves provide passive resistance which limits the degree of ozone exposure. The observed diminished genetic response is then likely a consequence of this reduced exposure.

Methods
Mandarin (Ottawa), PI 548379, and Fiskeby III, PI 438471, soybean cultivars were grown in a greenhouse under charcoal-filtered air conditions (<10 ppb ozone) for 25 days in October-November 2010 in the USDA-ARS greenhouse facilities in Raleigh, NC, USA. Plants were moved into continuously circulated tank reactors (CSTRs) in an adjacent greenhouse bay for a 3-day acclimation period prior to ozone exposure. CSTRs are cylindrical exposure chambers covered with Teflon film, designed for rapid mixing of gases [41, 42]. Ozone exposures began on the 28th day following planting using eight CSTRs divided into four experimental blocks, each block consisting of two CSTRs designated as either a low (25 ppb target) or high (75 ppb target) ozone treatment. Each CSTR contained one plant of each cultivar variety. Actual ozone concentrations measured during exposure were 27 ± 2 ppb and 74 ± 1 ppb in the low and high treatments, respectively. CSTR temperature, light level, and relative humidity averaged 34 ± 1 °C, 254 ± 13 µmol m⁻² s⁻¹ PAR, and 61 ± 1 %, respectively, during the exposure period.

For physiological measurements, plants were exposed to ozone as described above for 2 days prior to assessment. Leaf foliar injury was determined by estimating the percentage of adaxial leaf surface area exhibiting necrotic or chlorotic lesions and/or pigmented stipple on selected leaves [7]. Leaf air space was determined gravimetrically by weighing freshly harvested leaves before and after infiltration with water and reporting the difference in weight as a surrogate for the volume of air space within the leaf. Leaf specific weight was determined by removing six disks from each leaf using a cork borer of known area, drying the disks, and measuring the dry weight.

The fifth main stem trifoliate leaf (three leaves from each leaf stem) from the bottom of the plant was harvested from each genotype in one experimental block at each of 4 distinct time points from five to seven hours after exposure began at 0900, samples taken at 1400 and continued to be taken in 40 minute intervals. Leaf tissue was flash frozen in liquid nitrogen prior to transportation back to UNC-Charlotte on dry ice. Whole leaf tissue from each of the three leaflets from the 5th trifoliate leaf was pooled and extracted using the Qiagen Plant RNAeasy extraction kit (Qiagen, CA) followed by Ambion DNA-free treatment (Life Technologies, NY). RNA-seq libraries were subsequently created using Illumina TruSeq RNA-seq kit (Illumina, CA). Each library was quantified on the Bioanalyzer using RNA nano chips (Agilent, CA) prior to submission for sequencing at the David H. Murdock Research Institute Core Facility (Kannapolis, NC). Each library was run on a separate lane of a HiSeq2000, generating 100 base pair single end reads.

RNA-seq quality figures and statistics were produced by RSeQC, a package for assessing read and mapping quality, using the read_duplication script [43] (see Additional file 1: Figure. S1). The results of sequencing were mapped back to the Williams 82 soybean reference genome Glycine max Wm82a2v1 using TopHat [29]. Mapping was optimized for non-mammalian data sets as per program instructions by changing the maximum allowed intron size to 5Kb. The GFF file from the Williams82 reference genome was also provided in the mapping step. Differential gene and isoform analysis of the mapping results was determined using the Cufflinks cuffdiff program using the `-u` and `-b` options for individual time points, as well as across all four time points using the –T option. Differential expression was assessed by the cuffdiff program and corresponds to those genes returning a p < .05 as determined by cuffdiff. Analysis of the timing of differential gene expression across the four time points was accomplished using custom Python scripts (available upon request) and statistical measures from cuffdiff. Given that recent work by the SEQC/MACQ-III consortium finding that expression measures correlated well between RNA-seq and qPCR confirmation using qPCR was not done [44]. However, as part of ongoing analyses, targeted genes are being analyzed using qPCR. Functional analysis of expressed genes was determined using associated gene ontology (GO) terms for soybean gene models available on Soybase (www.soybase.org) and custom Python scripts [45]. Single enrichment analysis (SEA), a function of Agrigo was used to examine GO term enrichment in sample queries; results were confirmed using the Gene Ontologies distribution tool available at Soybase.org [45, 46]. SEA results return abundant GO term categories found in the sample set based on an adjusted p-value < .05.
CummeRbund, an extension of the cufflinks package for data visualization, was used for visualization of results and read dispersion [47] (see Additional file 1: Figure S1). Soybean orthologs of genes used in the Le Provost et al. study were identified using BLASTp [48].

Availability of supporting data

The data sets supporting the results of this article are available in the GEO repository. For review purposes, see http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=kfyzsiuudrbdqt&acc=GSE59076.

Additional file

Additional file 1: Figure S1. Read quality assessment using RSeQC. Here, number of identical reads (blue) is mapped against the number of reads mapped to identical locations (red) for each sample. Figure S2. Quality Assessment by cummeRbund estimating read overdispersion in the high ozone samples. Table S1. Gene identifiers and associated GO annotations for genes that related to oxidative or geninal stress response in ozone-tolerant soybean as shown in the Figure 4 heat map. Table S2. Putative soybean wax biosynthesis genes identified by BLAST of A. thaliana genes in Le Provost et al. and by querying the full set of G_max genes for those with biological process GO terms matching wax biosynthesis. Shown are the log2 fold change values of the FPKM results of the cufflinks analysis comparing Fiskeby Mandarin (ottawa), while negative fold change (red) denote higher expression in Fiskeby.

Abbreviations

GO: Gene Ontology; CSTR: Continuously circulated tank reactors; SAR: Systemic Acquired Resistance.

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

AW participated in sample collection, performed the data analysis and drafted the manuscript. JSheridan and SS preformed the RNA extraction and library construction. Gm a x

Acknowledgments

The authors thank Renee Tucker for assistance with plant growth and Jeff Barton for operating the ozone exposure system. Funding for Kent Burkey and Amy Burton provided by USDA-ARS CRIS# 6645–11000–008–00D and United Soybean Board grant #0286. Adam Whaley was supported by NSF Award #1100695 and Sajedeh Safari was supported by NSF Award #0822258. Funding for the genomics experiments was provided by UNC-Charlotte.

Author details

1 Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC 28223, USA. 2 United States Department of Agriculture, Agricultural Research Service, 3127 Ligon St, Raleigh, NC 27695, USA.

Received: 7 July 2014 Accepted: 18 May 2015 Published online: 04 June 2015

References

1. Prather M, Ehlhalt D, Dentener F, Derwent R, Drilgokonye C, Holland E, et al. Others: Atmospheric Chemistry and Greenhouse Gases, Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change JT Houghton, et al., 239–287. New York: Cambridge Univ. Press, 2001.

2. U.S. EPA. Air Quality Criteria for Ozone and Related Photochemical Oxidants (2006 Final). Washington, DC: U.S. Environmental Protection Agency; EPA/ 600/R-05/004F-CF, 2006.

3. Ainsworth EA, Yendrek CR, Sitch S, Collins WJ, Emerson LD. The effects of tropospheric ozone on net primary productivity and implications for climate change*. Annu Rev Plant Biol. 2012;63:637–61.

4. Mills G, Buse A, Gimeno B, Bermejo V, Holland M, Emerson L, et al. A synthesis of AOT40-based response functions and critical levels of ozone for agricultural and horticultural crops. Atmos Environ. 2007;41:2630–43.

5. Purcell LC, Specht JE. Physiological traits for ameliorating drought stress. In HR Boerema and JE Specht (ed). Soybean: Improvement, production, and uses. Third Edition. Agron. Monogr. 16. ASA, CSSA and SSSA Madison, WI p. 569–620.

6. Boeker M, Munttitering R, McGrath M, Burkey K, Decoteau D, Fuscus E, et al. The Ozone Component of Global Change: Potential Effects on Agricultural and Horticultural Plant Yield, Product Quality and Interactions with Invasive Species. J Integr Plant Biol. 2009;51:337–51.

7. Burkey KO, Carter JR TE. Foliar resistance to ozone injury in the genetic base of U.S. and Canadian soybean and prediction of resistance in descendant cultivars using coefficient of parentage. Field Crops Res. 2009;111:207–17.

8. Mauzerall DL, Wang X. PROTECTING AGRICULTURAL CROPS FROM THE EFFECTS OF TROPOSHERIC OZONE EXPOSURE: Reconciling Science and Standard Setting in the United States, Europe, and Asia. Annu Rev Energy Environ. 2001;26:237–68.

9. Wang X, Mauzerall DL. Characterizing distributions of surface ozone and its impact on grain production in China, Japan and South Korea: 1990 and 2010. Atmos Environ. 2004;38:4383–402.

10. Keen NT, Taylor OC. Ozone Injury in Soybeans Isoflavonoid Accumulation Is Related to Necrosis. Physiol Plant. 1975;55:731–3.

11. Heagle AS, Miller JE, Pursley WA. Influence of Ozone Stress on Soybean Response to Carbon Dioxide Enrichment: III. Yield and Seed Quality Crop Sci. 1998;38:128.

12. Bou Jaoudé M, Katerji N, Mastrovrii M, Rana G. Analysis of the ozone effect on soybean in the Mediterranean region: II. The consequences on growth, yield and water use efficiency. Eur J Agron. 2008;28:519–25.

13. Kettunen R, Overmyer K, Kangasjärvi J. The role of ethylene in the formation of cell damage during ozone stress. In Biology and Biotechnology of the Plant Hormone Ethylene II. Netherlands: Springer; 1999:299–305.

14. Tong D, Mathur R, Schere K, Kang D, Yu S. The use of air quality forecasts to assess impacts of air pollution on crops. Methodology and case study. Atmos Environ. 2007;41:8772–84.

15. Fishman J, Creilson JK, Parker PA, Ainsworth EA, Vining GG, Szarka J, et al. An investigation of widespread ozone damage to the soybean crop in the upper Midwest determined from ground-based and satellite measurements. Atmos Environ. 2010;44:2248–56.

16. Morgan PB, Mies TA, Boliero GA, Nelson RL, Long SP. Season-long elevation of ozone concentration to projected 2050 levels under fully open-air conditions substantially decreases the growth and production of soybean. New Phytol. 2006;170:333–43.

17. Kanofsky JR, Sima P. Single oxygen production in ozone-stressed subunits of ribulose bisphosphate carboxylase/oxygenase in ozone-stressed plants. Plant Physiol. 2001;125:1225–33.

18. Reich PB, Schoettle AW, Amundson RG. Effects of low concentrations of O3, leaf age and water stress on leaf diffusive conductance and water use efficiency in soybean. Physiol Plant. 1985;63:58–64.

19. Kudo A, Saji H, Tanaka K, Kondo N. Expression of Arabidopsis cysteolic ascorbate peroxidase gene in response to ozone or sulfur dioxide. Plant Mol Biol. 1995;29:479–89.

20. Conklin PL, Barth C. Ascorbic acid, a familiar small molecule intertwined in the response of plants to ozone, pathogens, and the onset of senescence. Plant Cell Environ. 2004;27:959–70.

21. Chen Z, Gallic DR. Increasing tolerance to ozone by elevating foliar ascorbic acid confers greater protection against ozone than increasing avoidance. Plant Physiol. 2005;138:1673–89.

22. Ernst D, Schaudner M, Langbeutels C, Sandermann Jr H. Ozone-induced changes of mRNA levels of β-1, 3-glucanase, chitinase and “pathogenness-related”-protein β in tobacco plants. Plant Mol Biol. 1992;20:673–82.

23. Reddy G, Arteca RN, Y-F DAI, Flores HE, Negm FB, Pell EJ. Changes in ethylene and polyamines in relation to mRNA levels of the large and small subunits of ribulose bisphosphate carboxylase/oxygenase in ozone-stressed potato foliage. Plant Cell Environ. 1993;16:819–26.

24. Runekles VC, Chevone BL. Crop responses to ozone. Surf Level Ozone Expo Their Eff Veg. 1992;1:89–270.
25. Overmyer K, Brosché M, Kangasjärvi J. Reactive oxygen species and hormonal control of cell death. Trends Plant Sci. 2003;8:335–42.

26. Maccarrone M, Veldink GA, Wiegenthart JF. Thermal injury and ozone stress affect soybean lipoygenases expression. FEBS Lett. 1992;309:225–30.

27. Galant A, Koester RP, Ainsworth EA, Hicks LM, Jett JM. From climate change to molecular response: redox proteomics of ozone-induced responses in soybean. New Phytol. 2012;194:220–9.

28. Frei M, Tanaka JP, Wissuwa M. Genotypic variation in tolerance to elevated ozone in rice: dissection of distinct genetic factors linked to tolerance mechanisms. J Exp Bot. 2008;59:3741–52.

29. Trapnell C, Pachter L, Salzberg SL. TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25:1105–11.

30. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Others: Genome sequence of the palaeopolyploid soybean. Nature. 2010;463:78–83.

31. Trapnell C, Roberts A, Goff L, Pertaea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7:562–78.

32. Mittler R. Oxidative stress, antioxidants and stress tolerance. Trends Plant Sci. 2002;7:405–10.

33. Crisosto CH, Retzlaff WA, William LE, DeJong TM, Zoffoli JP. Postharvest performance evaluation of plum (Prunus salicina Lindel., Casselman') fruit grown under three ozone concentrations. J Am Soc Hortic Sci. 1993;118:497–502.

34. Norton JS, I, E D: Effect of ozone on storage of cranberries. In Proceedings of the American Society for Horticultural Science. Volume 93. Amer soc horticultural science 701 North Saint Asaph Street, Alexandria, VA 22314-1998; 1968:792.

35. Elstner EF, Osswald W, Youngman RJ. Basic mechanisms of pigment bleaching and loss of structural resistance in spruce (Picea abies) needles: advances in phytomedical diagnostics. Experientia. 1985;1:591–7.

36. Hellgren LI, Carlsson AS, Selldén G, Sandelius AS. In situ leaf lipid metabolism in garden pea (pisum sativum L.) exposed to moderately enhanced level of ozone. J Exp Bot. 1995;46:221–30.

37. Shepherd T, Wynne Griffiths D. The effects of stress on plant cuticular waxes. New Phytol. 2006;171:469–99.

38. Zhao TH, Wang JL, Wang Y, Sun JW, Cao Y. Effects of reactive oxygen species metabolic system on soybean (Glycine max) under exogenous chitosan to ozone stress. Bull Environ Contam Toxicol. 2010;85:59–63.

39. Kunst L, Saumels L. Plant cuticles shine: advances in wax biosynthesis and export. Curr Opin Plant Biol. 2009;12:721–7.

40. Le Provost G, Domergue F, Lalancre C, Campos PR, Grosbois A, Bert D, et al. Soil water stress affects both cuticular wax content and cuticle-related gene expression in young saplings of maritime pine (Pinus pinaster Ait). BMC Plant Biol. 2013;13:95.

41. Rogers HH, Jeffries HE, Stahel EP, Heck WW, Ripperton LA, Witherspoon AM. Measuring Air Pollutant Uptake by Plants: A Direct Kinetic Technique. J Air Pollut Control Assoc. 1977;27:1192–7.

42. Heck WW, Philbeck R B. and Dunning JA (1978) A continuous stirred tank reactor (CSTR) system for exposing plants to gaseous air contaminants. Principles, specifications, construction, and operation. US Dept. Agr. Washington, DC: Agricultural Research Series 181. U.S. Govnt. Printing Office; pp 32.

43. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012;28:2184–5.

44. Goff L, Trapnell C, Kelley D: cummeRbund: Analysis, Exploration, Manipulation and Visualization of Cufflinks High-Throughput Sequencing Data. R Package Version 2011, 1. http://www.bioconductor.org/packages/release/bioc/html/cummeRbund.html.

45. Grant D, Nelson RT, Cannon SB, Shoemaker RC. SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Res. 2010;38 suppl 1:D843–6.

46. Du Z, Zhou X, Ling Y, Zhang Z, Su Z. agrigO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res. 2010;38 suppl 2:W64–70.

47. Goff L, Trapnell C, Kelley D: cummeRbund: Analysis, Exploration, Manipulation and Visualization of Cufflinks High-Throughput Sequencing Data. R Package Version 2011, 1. http://www.bioconductor.org/packages/release/bioc/html/cummeRbund.html.

48. Attiehl ST, fish W, Miller W, Myers DW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.