Novel in vitro diagnosis of equine allergies using a protein array and mathematical modelling approach: A proof of concept using insect bite hypersensitivity

E. Marti b, *, X. Wang a, N.N. Jambari a, C. Rhyner c, J. Olzhausen c, J.J. Pérez-Barea a, G.P. Figueredo d, M.J.C. Alcocer a

a School of Biosciences, University of Nottingham, Sutton Bonington campus, Loughborough LE12 5RD, UK
b Department of Clinical Research and Veterinary Public Health, University of Bern, Switzerland
c Swiss Institute of Allergy and Asthma Research, University of Zürich, Davos, Switzerland
d School of Computer Sciences, Advance Data Analysis Centre (ADAC), University of Nottingham, Jubilee Campus, Nottingham, UK

A R T I C L E   I N F O

Article history:
Received 17 April 2015
Received in revised form 18 June 2015
Accepted 25 June 2015

Keywords:
Protein array
Allergens
Chemometrics
IgE
Horse

A B S T R A C T

Insect bite hypersensitivity (IBH) is a seasonal recurrent skin allergy of horses caused by IgE-mediated reactions to allergens present in the saliva of biting insects of the genus Culicoides, and possibly also Simulium and Stomoxys species (Fadok and Greiner, 1990; Wilson et al., 2001; Baselgia et al., 2006; Hellberg et al., 2006).

The estimated prevalence of IBH worldwide is variable, ranging from 3% to 11.6% in the UK (McCaig, 1973), 37.7% in parts of Germany (Littlewood, 1998), 10–60% in areas of Queensland, Australia (Riek, 1954) and 71.4% in some regions of The Netherlands (van Grevenhof et al., 2007). Interestingly, IBH is not found in Iceland, where Culicoides insects are absent, yet horses exported from Iceland for more than two years to countries where Culicoides are present show a prevalence of IBH of 50%, while the prevalence of IBH in their progeny born in mainland Europe is <10% (Björnsdóttir et al., 2006; Hallförrdóttir and Larsen, 1991).

Several IgE-binding salivary protein allergens from Culicoides have now been identified (for a review see Schaffartzik et al., 2012) in C. nubeculosus (Schaffartzik et al., 2010, 2011), C. sonorensis (Langner et al., 2009) and C. obsoletus (van der Meide et al., 2013; Peeters et al., 2013). Interestingly, despite the distinctively geographical distribution of the different Culicoides species, cross-reactivity has been observed amongst different species (Hallförrdóttir et al., 1989; Anderson et al., 1993) but recent studies indicate that allergens from Culicoides species originating from the environment of the horses display higher IgE-binding than allergens from species only rarely present in this environment (van der Meide et al., 2012).

The clinical symptoms, together with a thorough medical history, taking other conditions that may lead to the characteristic pruritus of the disease into account, are still considered gold standard for the diagnosis of IBH. To confirm the clinical diagnosis,
methods using crude whole body extract preparations as allergens have been used with mixed degrees of success such as intradermal skin test (Sloet van Oldruitenborgh-Oosterbaan et al., 2009; Wagner et al., 2009), in vitro histamine release (Wagner et al., 2008) and sulfide leukotriene (sLT) release assays (Marti et al., 1999; Baselgia et al., 2006), and serological IgE tests (Frey et al., 2008; van der Meide et al., 2012). Recent studies indicate that the use of pure recombinant Culicoides proteins may result in an improvement of serological IgE tests for diagnosis of IBH (van der Meide et al., 2014; Peeters et al., 2013). However, because of the high number of Culicoides protein allergens relevant for IBH, component resolved diagnosis (CRD) using ELISA is cumbersome and expensive.

With the advances in robotics and computational technology, the ability to produce microarrays has provided the possibility to spot and monitor thousands of individual molecule populations on a miniaturised scale, on solid phase-based media of the size of a microscope slide. These protein spots and their interaction with specific molecules (e.g. immunoglobulins) can easily be tracked by techniques like fluorescence and advanced algorithms for image processing and pattern recognition. It is now well described and broadly accepted that beyond the comprehensive qualitative coverage, the diagnostic value of these tests with selected allergens can be similar to standard laboratory methods such as UniCAP, ELISA and immunoblot tests (Renault et al., 2011).

With the availability of a panel of recombinant Culicoides allergens, we demonstrate here that a complex protein array can be used for accurate diagnosis of IBH. Furthermore, we demonstrate that classification methods such as Partial Least Squares Discriminant Analysis (PLSDA), widely used by engineers and scientists in chemometrics-based research, can be a useful tool in the analysis of complex multivariate data allowing the production of predictive and testable mathematical models.

2. Materials and methods

2.1. Horses

A total of 133 horses comprising 63 non-affected (healthy) controls, 35 IBH-affected horses, 6 horses affected with recurrent airway obstruction (RAO) and 29 horses showing classic symptoms of recurrent urticaria were included in the study. The mean age of the horses was 11 years (range = 1–27). There were no significant differences between the groups except that the 6 RAO-affected horses were significantly older than the other horses (mean age = 19 years). The horses consisted of 56 females and 77 males, and again there were no significant differences in the gender distribution between the groups (Chi-Square = 0.6). The horses consisted of various breeds including Icelandic horses, Warmblood horses, Franches-Montagnes, Arabians, Andalusians and ponies.

In the first part of the study a subgroup of the above mentioned horses consisting of all 35 IBH-affected horses and of 37 healthy controls was analysed. In this group 24 out of the 35 IBH-affected horses and 25 out of the 37 non-affected horses belonged to the Icelandic breed. The IBH-affected horses had typical clinical signs of IBH and a clinical history of recurrent seasonal dermatitis affecting the dorsal and sometimes also the ventral midline. Furthermore, the IBH-horses were all positive in a cellular sLT release assay with C. nubeculosus extract, used to confirm the clinical diagnosis of IBH (Baselgia et al., 2006). The 37 healthy control horses had no clinical signs or history of skin diseases and were all living on the same farms where at least one of the IBH-affected horses was living, i.e. they were living in an environment where the causative insects for IBH (Culicoides spp.) were present. Furthermore, the healthy control horses had a sLT release with C. nubeculosus extract that was below the defined cut off (Baselgia et al., 2006).

In order to test the robustness of the test, in the second part of the study new microarray analyses were carried out. A larger group of randomly selected clinically non-affected horses and horses that were affected with conditions which are potentially IgE-mediated, including recurrent urticaria (Jose-Cuñilleras et al., 2001; Rüfenacht et al., 2005; Hinden et al., 2012) and recurrent airway obstruction (Jose-Cuñilleras et al., 2001) were analysed. The ability to discriminate IBH affected horses from other IgE-mediated disease background should demonstrate the robustness of the test.

2.2. IgE serum determination by protein microarray

A complex protein microarray containing extracts (n = 240) and pure proteins (n = 120) from a wide range of protein families from food (fruit, dairy, seeds), epithelium, pollen, fish, mollusca, fungi and insects representing in total 200 environmental and food species (Table S1) was assembled essentially as described previously (Wulfert et al., 2012; Renault et al., 2011). The extracts and purified proteins were obtained from commercial suppliers, produced in house and donated mainly by Dr. Maria Antonietta Ciardiello (Istituto di Biosciences and Bioresources, Napoli, Italy). Additionally for this study, a whole body C. nubeculosus extract available commercially (XPB681A2.5, Greer allergy Immunotherapy, Lenoir, USA) and a female C. nubeculosus thorax and head extract, made as described in Peeters et al. (2013), as well as recombinant Culicoides allergens were used. The recombinant Culicoides allergens originating from C. nubeculosus (Cul n 1–10) and C. obsoletus (Cul o 1 and Cul o 2) allergens were produced in Escherichia coli and purified as described previously (Schaffartzik et al., 2010, 2011; Peeters et al., 2013). IgE-binding to these Culicoides extracts and recombinant allergens were measured individually, but for the PLSDA analyses the same procedure used for all the other families of proteins contained in the array was employed, i.e. all Culicoides-specific data have been amalgamated and averaged. The purified protein solutions and extracts were normalised and spotted using a Marathon microarrayer (Arrayjet, Roslin, Scotland) into 16 pad nitrocellulose FAST slides (Whatman Schleicher & Schuell, Dassal, Germany) to a final density of 12,288 spots/slide. After blocking with 3% BSA (w/v) in PBS, the microarrays were hybridised with the horse sera diluted 1:2 in PBS containing 0.2% Tween-20 (PBST) and 2% BSA and surface bound allergen specific equine IgE was quantified using mouse anti-horse IgE mAb (Wilson et al., 2006) and fluorophore labelled rabbit anti-mouse IgG Ab (Dylight 649, Rockland Inc., USA) both diluted at 1:400 in 0.2% PBST containing 1% BSA. The slides were then washed three times in 0.05% PBST, followed by five times with purified water, and finally dried by centrifugation (MSE Mistral 3000i, Sanyo, UK) at 300 × g for 10 min at room temperature as previously described (Renault et al., 2011). Dried slides were scanned in a Genepix 4000B (Molecular devices, USA) with the settings 440 and 310 for the PMTs at 635 and 532 nm, respectively, and image processed in GenePix Pro software v6.0.1.27 (Axon Instruments). Control microarray results, captured on individual slides (1 out of 16 pads), consisting of all reagents except horse serum, were subtracted from the sample slides to eliminate non-specific binding and inherent autofluorescence of some proteins using dedicated programs developed in house running on Matlab (Version 9. The Mathworks Inc., USA) using an Excel link toolbox (Mathworks) and Dataset Object (Version 5.0, Eigenvector Research Inc., USA).

2.3. Data analysis

Multivariate data analysis was carried out using the PLSD Toolbox (Version 5.8.3, Eigenvector Research Inc., USA) with principal components analysis for data exploration/visualization and partial least square discriminant analysis software (PLSDA) for discriminant
analysis and as general classifier. PLSDA is a technique based upon Partial Least Squares regression method for constructing predictive models (Wold et al., 2001). Due to its ability to handle data in which variables far out-number the number of samples and in which many features are correlated (multicollinearity), PLSDA is especially suited for the types of analysis described here. Once the PLS model has been built, the influence of individual features may be captured with variable influence on the projection (VIP) derived from the PLS coefficients for the optimal set of components (Wold et al., 2001). A threshold of α > 1 was used. Most of the data was means-centred and scaled to unit standard deviation (autoscale). Internal cross-validation was employed to assess the number of latent variables (data trends) necessary to build models that are as concise as possible with minimal predictive error. All the classification results were confirmed by cross-validation where random subsets of the data were left out. Regression models were built with increasing numbers of latent variables and the prediction error established for the left out samples for each of the models. This procedure is repeated until all samples have been left out once and an average prediction/classification error per number of latent variables is established. The result is an estimation of the most appropriate number of latent variables (with lowest error) as well as an estimation of the prediction/classification error to be expected when applying the model to new data.

3. Results and discussion

Previously, we have shown that complex protein arrays from extracts can be quantifiable, reproducible and that different Ig classes can be monitored simultaneously. We then showed that in human patients IgE responses to extracts and pure proteins correlated well with other conventional techniques such as ELISA and UniCAP data (Renault et al., 2011). Therefore, the rationale behind the format employed here is to maximise the coverage with extracts but simultaneously maintaining the specificity of the single pure proteins when these proteins are available. This allows component-resolved diagnostics (CRD), i.e. defining to which specific allergenic molecules an individual is sensitised. In the human allergy field, CRD permits a distinction between genuine sensitisation (sensitisation to the primary original sensitising molecule) and cross-reactivity (IgE-binding to similar allergenic molecules from different sources). This improves the selection of allergens for specific immunotherapy and thus results in higher success rates (Canonica et al., 2013).

Based on the same principles, we have adapted the array to equine serum samples. The general profiling of the dataset produced in the complex array described here has not revealed any significant trend due to the clinical diagnosis or bias due to processing, slide number, position, day of processing, or operator when analysed by principal component analysis tools (data not shown).

In complex and comprehensive arrays, a direct regression analysis inevitably results in poor convergence and spurious results due to the large number of variables per sample. Aiming at avoiding these well described problems, a multi-parametric approach was employed here. In order to determine the importance of the different variables in the discrimination of the IBH samples from other clinical groups, a PLSDA classification analysis was performed. In this analysis, by encoding class labels as binary vectors indicating class membership, i.e. 1 or 0 for belonging or not to the “healthy” class and the same for “IBH” or other classes, PLSDA attempts to find factors (latent variables) which both capture variance and achieve correlation (maximise covariance) that are relevant for the class membership prediction (Barker and Rayens, 2003; Perez et al., 2009). Once the model has been built, the influences of individual variables are indicated by the VIP (Variable Influence on Projection). The variables can then be ranked in scores after the choice of appropriate threshold (Bryan et al., 2008).

The initial PLSDA classification using the first subgroup of horses, i.e. IBH diagnosed horses (n = 35) against their healthy control animals (n = 37), was very encouraging (Table 1). Whilst a high separation of the two groups was attained with sensitivity and specificity of 100%, these values only indicated fitting of the available data imposed by the classifier. In order to test the mathematical model produced, a procedure able to quantify the predictive robustness of the model, referred here as cross validation, was carried out. In this mathematical process by randomly removing samples (splits), re-classifying them and double checking how well they can fit the model, the actual accuracy of the model can be established. The specificity and sensitivity values for the cross validation were 0.733 and 0.867, respectively (Table 1). This suggests that the predictive model at this stage was good but that many insignificant variables were considered.

As shown in Fig. 1, the VIPs of the mathematical model produced for the discrimination of IBH affected against healthy horses mainly relied upon one major variable (Culicoides averaged extract + pure protein spots) and 31 other minor variables out of the 200 other variables present in the array. The VIP score of Culicoides was five times higher than the protein with the next highest VIP score. The identification of one variable with such a high VIP score is quite a unique result, considering that the samples are originated from an outbred population where most of the individuals live under various environmental conditions (i.e. in various stables and various places in Switzerland). This confirms the importance of Culicoides allergens in the pathogenesis of IBH as well as the clinical selection of the animals. Whether the minor variables identified in this first analysis have any clinical relevance needs to be determined in the future. IgE-binding to these proteins could possibly be due to cross-reactivity, for example, to serine proteases and amylases which are present in Culicoides as well as in other allergens, such as Blattella or Dermatophagoides, for example. In order to remove the background noise and improve the robustness of the predictive mathematical model, the main VIPs (n = 32) indicated in Fig. 1 were selected and exclusively used in a new modelling round. As expected and presented in Table 1, this enhanced the sensitivity and specificity of the predictive (cross validation) mathematical model. As this selection is mainly driven by one variable, using the average fluorescence of Culicoides proteins (recombiant + extracts) as the only variable, the sensitivity of the test drops slightly but remains a statistical significant classification method.

It is now widely accepted that better standardisations of allergen preparation can be obtained with pure proteins than with complex extracts. Hence, pure proteins are increasingly used for more specific diagnosis of allergies. The improvement in success rates for specific immunotherapy preconized by CRD in humans (Canonica et al., 2013) suggests that the use of arrays with pure allergens should also be the future in equine allergy. As we discussed in previous papers (Renault et al., 2011), the sheer reality of numbers are daunting. In the plant kingdom alone some important domesticated crops possess genome sizes that are many times larger than the human genome. Hence, even after taking into consideration the polyphylly nature and different polymorphisms present, it is unlikely that single protein arrays with total coverage will ever be built. Hence, the hybrid system we are presenting here seems the best we can master at this time. To illustrate this point and as shown in Fig. 2, each one of the IBH animals responded differently against the individual pure Culicoides proteins present in this array. In few cases as shown by Cul o 2, Cul n 4 and 5 a very selective response of IBH affected versus healthy control horses was observed and in others, such as Cul n 1, 7 and 9, a more generic response was observed. The Culicoides extracts also showed a generic response
Table 1
Partial least square discrimination analysis (PLSDA) statistical data analysis of the centered (autoscale) and cross validated (Venetian blind, 20 LV, 30 data split).

<table>
<thead>
<tr>
<th></th>
<th>Before VIP selection</th>
<th>After VIP selection</th>
<th>Only culicoides (extracts + pure recombinant proteins)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Calibration</td>
<td>Cross validation</td>
<td>Calibration</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>1.000</td>
<td>0.733</td>
<td>1.000</td>
</tr>
<tr>
<td>Specificity</td>
<td>1.000</td>
<td>0.867</td>
<td>0.933</td>
</tr>
<tr>
<td>Error</td>
<td>RMSEC = 0.150</td>
<td>RMSECV = 0.440</td>
<td>RMSEC = 0.296</td>
</tr>
</tbody>
</table>

albeit the background levels of the controls are a case for concern. Similarly, high background levels have also been observed with Cul n 9. Although the two extracts statistically have the same effect on the diagnosis (ANOVA interaction = 0.352) the discrimination (Fig. 2C) observed using the thorax and head extract made from female C. nubeculosus (p = 0.0017) against the commercial whole body extract (Greer) (p = 0.0044) shows some marginal improvement. This difference might be due to an enrichment of the relevant allergens through removal of the males and of the abdomens, as salivary glands are located in the thorax (Wilson et al., 2008) and males have different salivary glands than females. The preparation of female thorax extract, however, is cumbersome and could not be scaled up for a commercial test. Variations as described above are often encountered in complex protein preparations. Thus, protein extracts alone might not always produce a reliable diagnosis.

Hence, the results so far suggest that at the present stage, the use of an aggregated response of extracts and pure proteins for diagnostics should be considered. For immunotherapy purposes and specifically for IBH-affected horses, the use of additional recombinant proteins, as the recently identified and expressed C. obsoletus proteins (van der Meide et al., 2013) will be desirable towards an improved specific patient tailored immunotherapy of IBH and will thus need to be included on future arrays.

Although ideal, mathematical models for diagnosis of diseases using calibrated groups such the ones presented thus far are a long way from the reality of the clinical laboratories worldwide. Hence, in order to test the robustness of the diagnostic test another batch of sera from animals clinically diagnosed with RAO (n = 6) and urticaria (n = 29), i.e. with presumptive high IgE titres, and additional healthy controls were reanalysed using the same complex protein array.

Again the initial principal component analysis was unable to distinguish significant operational bias on the dataset. By using the averaged value of the extract plus pure proteins of the 200 families described in Table S1, the PLSDA of IBH horse against the others (healthy + RAO + urticaria) resulted in a sensitivity of 0.914 and a specificity of 0.939 for the calibration and of 0.856 and 0.908, respectively, for the cross validation after optimisation. Interestingly, the VIPs for this classification produced again the unique pattern of Culicoides proteins (Fig. 3). Using this approach, the predicted probability of belonging to the IBH group can be calculated and is shown in Fig. 4. With the exception of one animal (45% probability of IBH) the remaining animals showed >70% probability of belonging to this group, or in other words to be diagnosed as IBH.

The direct use of four simultaneous class groups (IBH, healthy, RAO, urticaria) produced no significant results, suggesting that for each specific disease the calibrated set of variables needs to be established and the mathematical model optimised. Thus, every patient sample will first go through the different previously calibrated mathematical routines that determine the probability of

Fig. 1. Variable influence on projection (VIP) calculated by partial least square discrimination analysis software (PLSDA) from 35 IBH affected and 37 healthy control horses using 200 averaged variables (extracts + pure proteins). The averaged Culicoides nubeculosus and obsoletus (extract + pure proteins) variable was the major variable identified as important for this classification. Some minor contributing variables are also shown.
Fig. 2. IgE response in fluorescence arbitrary units (FAU) of the 37 healthy controls (left, in white triangles) and 35 IBH-affected horses (right, black star) toward representatives of specific recombinant Culicoides allergens and Culicoides extracts. (A) Recombinant Culicoides nubeculosus allergens, (B) recombinant Culicoides obsoletus allergens, (C) female Culicoides nubeculosus thorax and head extract (black bars) and Culicoides nubeculosus whole body extract (Greer; grey bars).
Fig. 3. Variable influence on projection (VIP) calculated by partial least square discrimination analysis software (PLSDA) from IBH affected (n = 35), healthy controls (n = 63), RAO (n = 6) and urticaria (n = 29) horses using 200 variables. The averaged Culicoides nubeculosus and obsOLEtus (extract + pure proteins) variable again was the major variable identified as important for this classification. Some minor contributing variables are also shown.

Fig. 4. Prediction probability of belonging to the IBH group of horses. Using the calibration parameters shown in figure 3, the predicted probability of each horse to belong to the IBH affected group (n = 37) when compared to control horses (healthy + RAO + Urticaria; n = 98) is depicted.

belonging to a clinically defined group. This will allow a more meaningful interpretation of the numeric data readout, i.e. of the sensitisation pattern of each patient. In our laboratory pre calibrated routines are now being setup for other allergic diseases, using samples from clinically well characterised affected and control horses.

Interestingly and as expected, the classification of the urticaria group against the others showed a much more complex set of VIPs involving mainly fungal and pollen proteins. A full description, analysis and interpretation of the urticaria results should follow the publication of this article.

4. Conclusions

The profiling array technique in this work has shown that protein microarrays containing complex extracts and pure proteins can be used for the diagnosis of allergic diseases in horses. Besides the obvious advantages such as general profiling and use of few microliters of samples, this technique permits automation and generation of mathematical predictive models that can support the clinical diagnosis of allergic diseases. For IBH, the sensitivity and specificity values obtained confirmed the high discrimination
power of the complex extracts plus pure recombinant Cuculoides proteins on this allergic disease and made possible attaining a robust predictive mathematical model. It also further demonstrates the specificity of the protein array method for identification of specific IgE sensitisations in allergic diseases when the causative allergens are included on the array.

Conflict of interest

None.

Acknowledgements

We thank Professor Petra Roosje, Clinical Dermatology Unit, Department of Clinical Veterinary Medicine, University of Bern, for providing serum samples from horses affected with recurrent urticaria. The cooperation of many horse owners is gratefully appreciated. This work was supported by the Swiss National Science Foundation grant no. 310030-160196/1 and by Hermes Fellowships programme at the University of Nottingham grant no 13b/1. X. Wang was partially supported by PetPlan Charitable Trust.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.vetimm.2015.06.013

References


Perez, N.F., Ferre, J., Boque, R., 2009. Calculation of the reliability of classif...