CrossMark

ORIGINAL ARTICLE

Use of RSM and CHAID data mining algorithm for predicting mineral nutrition of hazelnut

Meleksen Akin[1] · Ecevit Eyduran[2] · Barbara M. Reed[3]

Abstract Defining optimal mineral-salt concentrations for in vitro plant development is challenging, due to the many chemical interactions in growth media and genotype variability among plants. Statistical approaches that are easier to interpret are needed to make optimization processes practical. Response Surface Methodology (RSM) and the Chi-Squared Automatic Interaction Detection (CHAID) data mining algorithm were used to analyze the growth of shoots in a hazelnut tissue-culture medium optimization experiment. Driver and Kuniyuki Walnut medium (DKW) salts (NH_4NO_3, $Ca(NO_3)_2 \cdot 4H_2O$, $CaCl_2 \cdot 2H_2O$, $MgSO_4 \cdot 7H_2O$, KH_2PO_4 and K_2SO_4) were varied from 0.5× to 3× DKW concentrations with 42 combinations in a IV-optimal design. Shoot quality, shoot length, multiplication and callus formation were evaluated and analyzed using the two methods. Both analyses indicated that NH_4NO_3 was a predominant nutrient factor. RSM projected that low NH_4NO_3 and high KH_2PO_4 concentrations were significant for quality, shoot length, multiplication and callus formation in some of the hazelnut genotypes. CHAID analysis indicated that NH_4NO_3 at ≤1.701× DKW and KH_2PO_4 at >2.012× DKW were the most critical factors for shoot quality. NH_4NO_3 at ≤0.5× DKW and $Ca(NO_3)_2$ at ≤1.725× DKW were essential for good multiplication. RSM results were genotype dependent while CHAID included genotype as a factor in the analysis, allowing development of a common medium rather than several genotype specific media. Overall, CHAID results were more specific and easier to interpret than RSM graphs. The optimal growth medium for *Corylus avellana* L. cultivars should include: 0.5× NH_4NO_3, 3× KH_2PO_4, 1.5× $Ca(NO_3)_2$.

Introduction

Growth medium salts, plant growth regulators, temperature and lighting are all key factors for improving in vitro plant growth. Tissue culture medium optimization studies have traditionally focused on a few factors studied at the same time, and based on simple ANOVA analysis or classical factorial designs. Factorial designs require a large number of treatments, even when only a few factors are included (Compton and Mize 1999; Ibañez et al. 2003; Mize et al. 1999; Nas et al. 2005). Experimental designs and statistical analyses that are able to evaluate the effect of many factors with various levels and their interactions on mineral nutrition of in-vitro plants are required for better optimization process.

New methodologies for improving in vitro shoot growth by changing mineral nutrients include using advanced statistical models such as response surface methodology (RSM) and neuro-fuzzy logic (Alanagh et al. 2014; Gago et al. 2011; Niedz and Evens 2007). RSM is a statistical technique

✉ Meleksen Akin
akinmeleksen@gmail.com

1 Department of Horticulture, Agricultural Faculty, Igdir University, Igdir, Turkey

2 Biometry Genetics Unit, Department of Animal Science, Agricultural Faculty, Igdir University, Igdir, Turkey

3 USDA-ARS-Retired, National Clonal Germplasm Repository, 33447 Peoria Rd, Corvallis, OR 97333, USA

🖄 Springer

for modeling and analysis of responses affected by several factors, with the main objective to optimize the response. If the response is a linear function of the factors, it is a first order RSM model (multiple regression). If curvature is present in the response surface, higher degree of polynomial should be used, for example second order polynomial. The parameters within the polynomials are estimated according to the Method of Least Square. Response Surface Methods are designs and models which allow optimization of continuous factors (Bradley 2007; Montgomery 2005). Therefore, the main drawback with respect to tissue culture medium optimization is that nominal variables such as genotype cannot be included as a factor in the analysis. RSM simultaneously evaluates polynomial relationships between several independent variables and the dependent variable, and provides genotype specific models showing general areas of optimal plant growth. Computer generated RSM optimal designs allow testing of multiple nutrient factors at once, reducing treatments, time, labor and explant number compared to traditional factorial designs (Anderson and Whitcomb 2005; Hand and Reed 2014; Reed et al. 2013). Other statistical techniques used for better understanding mineral nutrition requirements are artificial neural network and neurofuzzy logic algorithms. The neurofuzzy logic algorithm is a hybrid technology that combines artificial neural network and fuzzy logic techniques. The algorithm utilizes IF (condition) and THEN (observed behavior) rules to model non-linear relationships between several independent variables and responses (Gago et al. 2011).

Another option for data analysis that has not been used for plant tissue culture optimization is the Chi-Squared Automatic Interaction Detection (CHAID) data mining algorithm. As opposed to regression analysis, CHAID is a non-parametric technique, and therefore does not require distributional assumptions like normality and linearity. The algorithm can project interactions between variables, and non-linear effects which are generally missed by traditional statistical techniques (Hébert et al. 2006). The tree-based algorithm applied to the data obtained from an optimal design would help to more precisely define the optimum concentrations of salts in tissue culture media and to better understand any interactions. CHAID allows the analysis of responses of several genotypes simultaneously, which could contribute to developing a common tissue-culture medium rather than several genotype-specific media.

CHAID constructs a visual (non-binary) decision tree that contains many homogenous subsets with multi-way node splitting from a heterogeneous data set, by selecting explanatory variables that significantly interact with a response variable (Ali et al. 2015). The algorithm is appropriate for nominal, ordinal, and continuous data. Decision tree building processes consist of merging, splitting and stopping steps. Optimal splits are identified by Chi square statistics when the dependent variable is categorical, and F-test is used when the response is continuous. Continuous variables are split into categories with similar number of observations. CHAID splits the explanatory variables based on their smallest Bonferroni adjusted p-value. The categories of the independent variables are used for calculating p-values to check whether the mean responses are same for different categories of the predictors. If the p-values are not significant the pair is merged into a single group. A group with three or more categories is evaluated to define the most significant binary split. The splitting process continues until the node size is less than the predefined minimum node size value. The decision tree components are root node (containing the dependent variable), parent node (the first two or more categories after the data is split), child nodes (explanatory variable categories below the parent categories) and the terminal node is the last category. The most significant category on the dependent variable is at the top of the tree and the least important one (terminal node) is at the bottom (Rashidi et al. 2014; Statistics-Solutions 2016).

This study was designed to compare the conclusions of RSM and CHAID analysis for in-vitro culture data, and to provide practical approaches for tissue culture medium optimization. Optimal shoot quality, shoot length, multiplication, and callus formation were determined with each statistical technique.

Materials and methods

Plant material and in vitro culture conditions

Shoots of *C. avellana* L. hazelnuts 'Dorris', 'Wepster' and 'Zeta' were micropropagated on COR-2013 medium (Hand et al. 2014; Hand and Reed 2014) consisting of modified DKW (Driver and Kuniyuki 1984) medium salts [1.5× Ca(NO$_3$)$_2$·4H$_2$O, 1.5× MgSO$_4$·7H$_2$O and 1.5× KH$_2$PO$_4$; 4× H$_3$BO$_3$, 4× Na$_2$MoO$_4$·2H$_2$O, 4× Zn(NO$_3$)$_2$·6H$_2$O, 0.5× MnSO$_4$·H$_2$O, 0.5× CuSO$_4$·5H$_2$O] with 30 g L^{-1} glucose, 200 mg. L^{-1} sequestrene 138 Fe EDDHA, 2 mg L^{-1} thiamine, 2 mg nicotinic acid L^{-1}, 2 mg. L^{-1} glycine, 1 g L^{-1} myo-inositol, 50 mg N^{6}benzyladenine (BA), and 0.5% (w/v) agar (PhytoTechnology Laboratories A1111). Each vessel (Magenta GA7, Magenta, Chicago, IL) held 40 ml medium, and was autoclaved for 20 min at 121 °C. The growth room conditions consisted of 80 µmol m^{2}s^{-1} light intensity with a 16-h photoperiod of half warm-white and half cool-white fluorescent lamps (Ecolux Starcoat, 32 W; GE, Fairfield, CT) at 25 ± 2 °C.

A RSM IV-optimal experimental design was set up by Design-Expert software (Design-Expert 2010) using the standard salt composition of DKW medium (1×) with the modified micronutrients listed above (Hand and Reed

2014). A six factor design with NH_4NO_3, $Ca(NO_3)_2 \cdot 4H_2O$, $CaCl_2 \cdot 2H_2O$, $MgSO_4 \cdot 7H_2O$, KH_2PO_4 and K_2SO_4 salts were tested in a range of 0.5–3.0× DKW. Forty-two treatments were assigned with the last two treatments as controls of DKW salts with modified micronutrients (Table 1).

Shoots were cut to 3 cm with the apical meristem removed. During the first transfer the callus was removed and shoots were transferred to new medium. In the second transfer shoots were reduced to 3 cm by removing the base and the apical meristem. For each treatment two boxes with five shoots for each cultivar were used (n = 10). Boxes were randomized on the growth room shelf. Shoots were grown on each treatment medium for 10 week with the first and second transfers for 3 week, and the last transfer for 4 week.

Data

Three shoots from each culture vessel (n = 6) were evaluated as follows: shoot quality, a subjective visual assessment of shoot vigor and form was evaluated as 1 = poor, 2 = moderate and 3 = good. Shoots longer than 5 mm were counted. The longest shoots were measured in millimeters. Callus formation was evaluated as: 1 = callus ≥2 mm, 2 = callus <2 mm, and 3 = absent (no callus present).

Statistical analysis

The mean response (quality, shoot length, shoot number and callus) of six shoots per treatment for each genotype was assessed by RSM using Design Expert 8 software (Design-Expert 2010). The factors and their polynomial relationships and interactions which affected plant growth responses were determined at $p < 0.05$. RSM does not provide an option for defining a variable structure, all factors are treated as continuous. Graphical models of RSM were produced based on the six salt-factor design space. Genotype was not included as a factor. Separate models were built for each genotype. The factors with highest statistical impact were assigned as axes of the graphs.

CHAID analyzed quality, shoot length, shoot number and callus dependent variables based on the six basal salts and the three genotypes ('Dorris', 'Wepster', and 'Zeta') independent variables. The mean response of six shoots per treatment was calculated for each genotype. Therefore, the dependent variables and the basal salts were treated as continuous variables, only genotype was selected as a nominal variable within the algorithm. To prune the redundant structuring of the tree diagram, the ideal minimum plant numbers for parent and child nodes were assigned based on the best tree diagram with highest Pearson correlation and no overlapping of the factors. The minimum plant numbers for quality were assigned as 28:14, shoot length 6:3, shoot number 24:12 and callus 12:6. Statistical analyses of

the tree-based CHAID algorithm were performed with IBM SPSS Statistics for Windows (Version 22.0, Armonk, NY) software (SPSS 2013).

Results and discussion

This study was designed to compare statistical methods for improving, and making more practical, the optimization process for mineral nutrients of tissue culture medium using RSM and CHAID data mining methodologies. RSM is a computer assisted technique which models optimal growth areas using selected treatment points within the range of all possible treatment combinations, and therefore reduces the number of treatments required for traditional factorial designs (Anderson and Whitcomb 2005). CHAID is a data mining algorithm used for constructing decision trees with homogenous sub-groups. It is useful for detecting non-linear and interaction effects without requiring linearity and normality assumptions (Hébert et al. 2006).

Quality

Plant quality rating is a subjective evaluation of general plant growth consisting of leaf and shoot health, as well as multiplication (Niedz et al. 2007). The RSM models for quality were significant ($p < 0.05$) for all of the genotypes tested (Table 2). Compared to DKW medium, increased KH_2PO_4 and lower K_2SO_4 concentrations affected the quality of all three cultivars ($p < 0.05$) and the NH_4NO_3 requirement was also low (Figs. 1, 2). For 'Dorris' there were interactions ($p < 0.05$) of several factors that impacted the shoot quality (Table 2). 'Dorris' required very high KH_2PO_4 and low K_2SO_4 concentrations and moderate to high $Ca(NO_3)_2$ for higher quality shoots (Fig. 1a). Low NH_4NO_3 was a significant factor for quality of 'Wepster', as were very high KH_2PO_4 and low K_2SO_4 concentrations (Fig. 1b). 'Zeta' quality models indicated that low to medium KH_2PO_4, low K_2SO_4 and low NH_4NO_3 were all required for the best growth. Improved growth was seen with several of the treatment combinations (Fig. 2).

The CHAID data mining algorithm indicated that the most significant factor for plant quality was NH_4NO_3, followed by KH_2PO_4 and K_2SO_4 (Fig. 3). The genotype effect in the CHAID algorithm was determined to be an insignificant source of variation for plant quality. The Pearson correlation coefficient between actual and predicted values for quality was 0.661 ($p < 0.01$), indicating a medium to high predictive ability of the algorithm. All the plants used in the analysis (Node 0) were split into two nodes (Nodes 1 and 2) based on the response to NH_4NO_3 concentration. The decisive cut-off value of NH_4NO_3 was 1.701×. The group of shoots with the best response to NH_4NO_3 ≤1.701× (Node 1) was

Table 1 Six factor design including 42 treatment points

| Treatments	Factor 1 NH_4NO_3	Factor 2 $Ca(NO_3)_2 \cdot 4H_2O$	Factor 3 $CaCl_2 \cdot 2H_2O$	Factor 4 $MgSO_4 \cdot 7H_2O$	Factor 5 KH_2PO_4	Factor 6 K_2SO_4
1	0.50	3.00	3.00	3.00	3.00	0.50
2	0.50	0.93	3.00	3.00	0.50	0.50
3	3.00	2.65	2.30	3.00	1.28	2.19
4	3.00	0.50	3.00	1.70	3.00	0.51
5	3.00	2.65	2.30	3.00	1.28	2.19
6	0.50	1.71	3.00	0.50	3.00	3.00
7	0.50	3.00	3.00	0.93	0.50	0.50
8	1.75	0.50	0.50	1.74	0.50	0.50
9	3.00	3.00	0.50	1.73	0.50	1.77
10	0.50	1.71	3.00	0.50	3.00	3.00
11	1.81	3.00	0.50	0.50	3.00	0.50
12	3.00	0.50	1.75	3.00	0.50	1.76
13	1.76	1.73	0.50	3.00	0.50	3.00
14	0.50	0.50	0.50	3.00	3.00	1.80
15	0.50	3.00	3.00	3.00	3.00	0.50
16	3.00	3.00	3.00	0.50	2.00	1.75
17	3.00	1.75	0.50	3.00	2.00	0.50
18	0.50	3.00	0.50	1.76	2.00	3.00
19	1.70	1.79	1.79	1.80	2.13	1.71
20	1.70	1.79	1.79	1.80	2.13	1.71
21	0.50	3.00	3.00	3.00	0.50	2.58
22	1.70	1.80	1.80	1.80	2.08	1.70
23	3.00	3.00	1.70	3.00	3.00	3.00
24	0.50	1.00	3.00	2.50	2.75	0.63
25	0.50	3.00	0.93	3.00	0.50	0.50
26	0.50	0.50	1.75	0.50	2.01	0.50
27	0.50	1.74	0.50	0.50	0.50	1.75
28	1.75	0.50	3.00	0.50	0.50	1.77
29	0.75	1.30	0.50	3.00	1.21	1.30
30	3.00	0.50	0.50	0.50	2.70	3.00
31	1.70	1.80	1.80	1.80	2.08	1.70
32	3.00	1.74	1.75	0.50	0.50	0.50
33	3.00	1.75	3.00	1.74	0.50	3.00
34	1.75	3.00	1.74	0.50	0.50	3.00
35	0.95	3.00	0.50	2.58	0.80	2.10
36	0.50	3.00	3.00	0.95	3.00	1.18
37	1.70	1.80	1.80	1.80	2.08	1.70
38	1.75	0.50	3.00	3.00	2.01	3.00
39	0.50	0.50	1.74	1.73	0.50	3.00
40	2.60	3.00	3.00	3.00	0.50	0.50
41	1.00	1.00	1.00	1.00	1.00	1.00
42	1.00	1.00	1.00	1.00	1.00	1.00

Design points 1–40 for investigating the effects of individual factors on mineral nutrition of hazelnut cultivars and DKW medium controls (points 41–42). DKW medium concentrations 1×: NH_4NO_3 (1416 mg), $Ca(NO_3)_2 \cdot 4H_2O$ (1960 mg), $CaCl_2 \cdot 2H_2O$ (147 mg), $MgSO_4 \cdot 7H_2O$ (740 mg), KH_2PO_4 (259 mg), K_2SO_4 (1560 mg). All treatments included modified minor nutrients (Hand and Reed 2014)

Table 2 DKW nutrient factors that had significant effects on four growth responses for each hazelnut cultivar at *p-value* (<0.05)

Cultivars	Measured responses			
	Quality	Shoot length	Shoot number	Callus
Dorris	Model (0.0217) KH_2PO_4 (0.0029) K_2SO_4 (0.0412) $NH_4NO_3 \times Ca(NO_3)_2 \cdot 4H_2O$ (0.0020) $Ca(NO_3)_2 \cdot 4H_2O \times MgSO_4 \cdot 7H_2O$ (0.0331)	Model (0.0255) NH_4NO_3 (0.0021) KH_2PO_4 (0.0008) $NH_4NO_3 \times Ca(NO_3)_2 \cdot 4H_2O$ (0.0157) $Ca(NO_3)_2 \cdot 4H_2O \times MgSO_4 \cdot 7H_2O$ (0.0142)	Model (0.0026) NH_4NO_3 (0.0004) $Ca(NO_3)_2 \cdot 4H_2O$ (0.0001) $MgSO_4 \cdot 7H_2O$ (0.0275) KH_2PO_4 (0.0013) $NH_4NO_3 \times K_2SO_4$ (0.0348) $MgSO_4 \cdot 7H_2O \times K_2SO_4$ (0.0135)	Model (NS[a])
Wepster	Model (0.0070) NH_4NO_3 (<0.0001) KH_2PO_4 (0.0011) K_2SO_4 (0.0171)	Model (0.0070) NH_4NO_3 (0.0001) KH_2PO_4 (0.0003)	Model (0.0015) NH_4NO_3 (0.0022) K_2SO_4 (0.0165) $NH_4NO_3 \times CaCl_2 \cdot 2H_2O$ (0.0004)	Model (0.0195) NH_4NO_3 (<0.0001) $NH_4NO_3 \times K_2SO_4$ (0.0417)
Zeta	Model (0.0153) NH_4NO_3 (0.002) KH_2PO_4 (0.0187) K_2SO_4 (0.0067)	Model (0.003) NH_4NO_3 (<0.0001) K_2SO_4 (0.0342)	Model (0.0006) NH_4NO_3 (<0.0001) $Ca(NO_3)_2 \cdot 4H_2O$ (<0.0001)	Model (0.014) NH_4NO_3 (0.0002) KH_2PO_4 (0.0399) K_2SO_4 (0.0139) $NH_4NO_3 \times MgSO_4 \cdot 7H_2O$ (0.0344)

[a] *NS* not significant

Fig. 1 Response surface graph of mineral nutrient effects on hazelnut shoot quality for **a** 'Dorris', **b** 'Wepster' and **c** 'Zeta'. The quality ratings were 1 = poor, 2 = moderate, 3 = good and highest (*red-yellow*) to lowest quality (*green–blue*). (Color figure online)
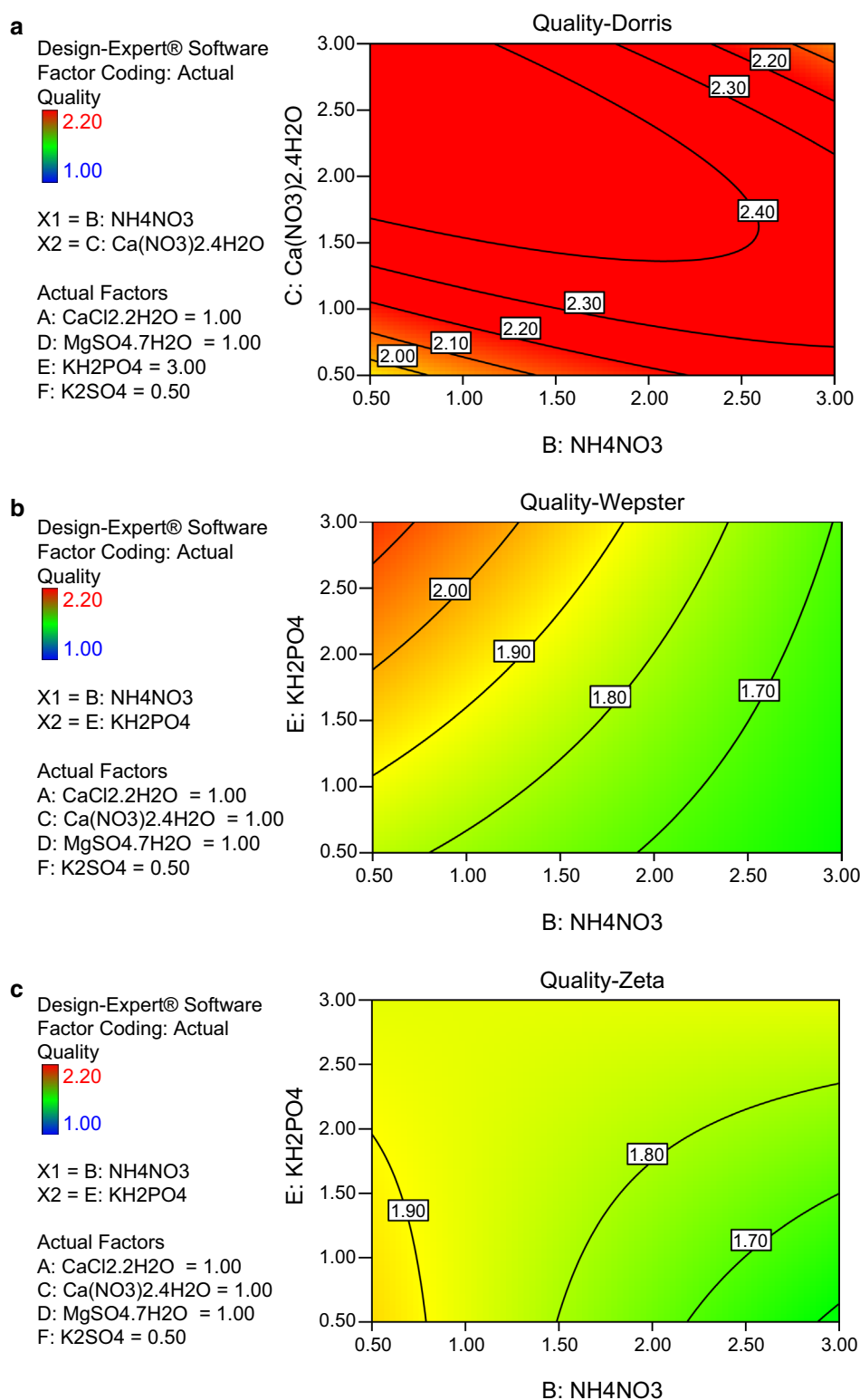
statistically different (adj. p < 0.001) from the group of shoots responding best at $NH_4NO_3 > 1.701\times$ (Node 2). The quality of shoots grown on $NH_4NO_3 \leq 1.701\times$ was also influenced by KH_2PO_4, while the quality of shoots on $NH_4NO_3 > 1.701\times$ was significantly affected by K_2SO_4 (adj. p < 0.001). Node

4, the group of plants exposed to $NH_4NO_3 \leq 1.701\times$ and $KH_2PO_4 > 2.012\times$ had significantly higher quality (2.037) than the other nodes (Fig. 3). This indicates that the resulting optimal medium for overall shoot quality would be $NH_4NO_3 \leq 1.701\times$, $KH_2PO_4 > 2.012\times$ for all three genotypes.
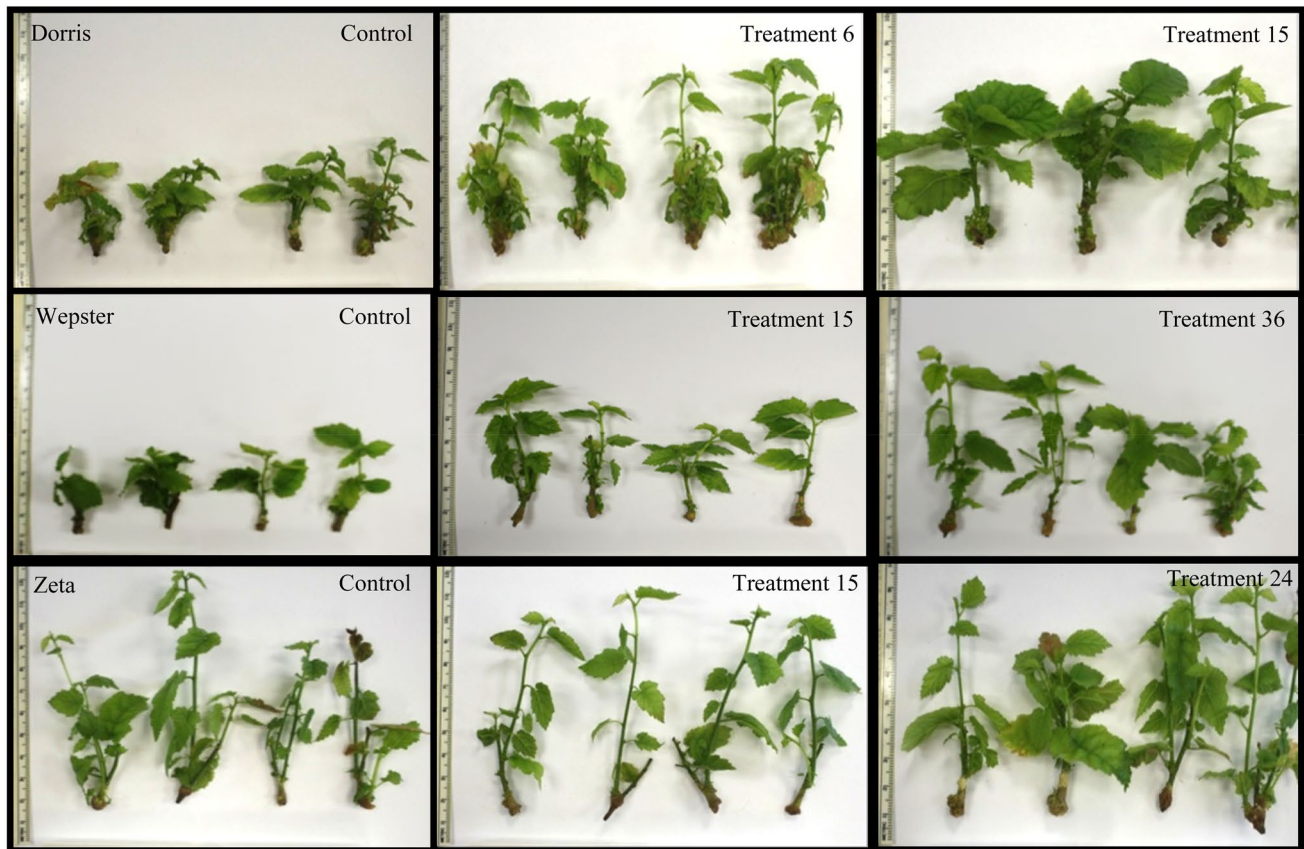
Fig. 2 Shoots of **a** 'Dorris', **b** 'Wepster' and **c** 'Zeta' grown on the control DKW salts and two treatments which produced higher plant quality

The differences in KH_2PO_4 and NH_4NO_3 requirement of plants noted in earlier studies could be related to genotype, the statistical methods applied, and the concentration ranges of variables used (Hand et al. 2014; Nas and Read 2004; Reed et al. 2013). Hand et al. (2014) linked quality with nitrogen factors, mesos and potassium sulfate, but the response was greatly dependent on genotype. All of the hazelnut genotypes in the current study required very low K_2SO_4 for better quality according to RSM graphs. K_2SO_4 was also significant in CHAID but only when NH_4NO_3 $\leq 1.701 \times$. $Ca(NO_3)_2$ was significant for quality of 'Dorris' in RSM, but wasn't significant in CHAID (Fig. 3). This difference could be due to the fact that CHAID algorithm simultaneously evaluated all three hazelnut cultivars in the present study. The cultivars were analyzed separately with RSM because genotype is not a continuous variable. Genotype was also evaluated separately in previous research using RSM (Hand et al. 2014; Reed et al. 2013). RSM is generally used to define the relationships between continuous (measurable) independent and dependent (response) variables, whereas CHAID is able to deal with categorical as well as continuous variables, and this allows evaluation of genotype as a factor, which could help to develop a general medium for diverse hazelnuts.

Shoot length

The RSM models for shoot length were statistically significant for all the hazelnut cultivars ($p < 0.05$) (Table 2). Interactions of several factors influenced ($p < 0.05$) shoot length of 'Dorris' in RSM (Table 2). 'Dorris' required high KH_2PO_4 and $Ca(NO_3)_2$, but low NH_4NO_3 for an ideal shoot length of 40 mm (Fig. 4a). 'Wepster' had the same requirements of KH_2PO_4 and NH_4NO_3 for longer shoots (40 mm), but did not require high $Ca(NO_3)_2$ concentrations (Fig. 4b). Shoots of 'Zeta' were typically long and the low-medium NH_4NO_3 and low-medium K_2SO_4 concentrations found in DKW were all that needed for good (40–50 mm) shoot length (Fig. 4c).

A CHAID decision tree diagram was constructed to predict shoot length from several significant factors: genotype, NH_4NO_3, KH_2PO_4, $CaCl_2$, and K_2SO_4 (Fig. 5). There was a very high Pearson correlation of 0.853 between actual and predicted shoot length values ($p < 0.01$). The main factor affecting shoot length was NH_4NO_3 (adj. $p < 0.001$), and the second most important factor was genotype (adj. $p < 0.001$). KH_2PO_4, $CaCl_2$, and K_2SO_4 were also influential for shoot length (adj. $p < 0.001$). All of the plants in Node 0 at the top of the tree diagram were split into two nodes (Nodes 1
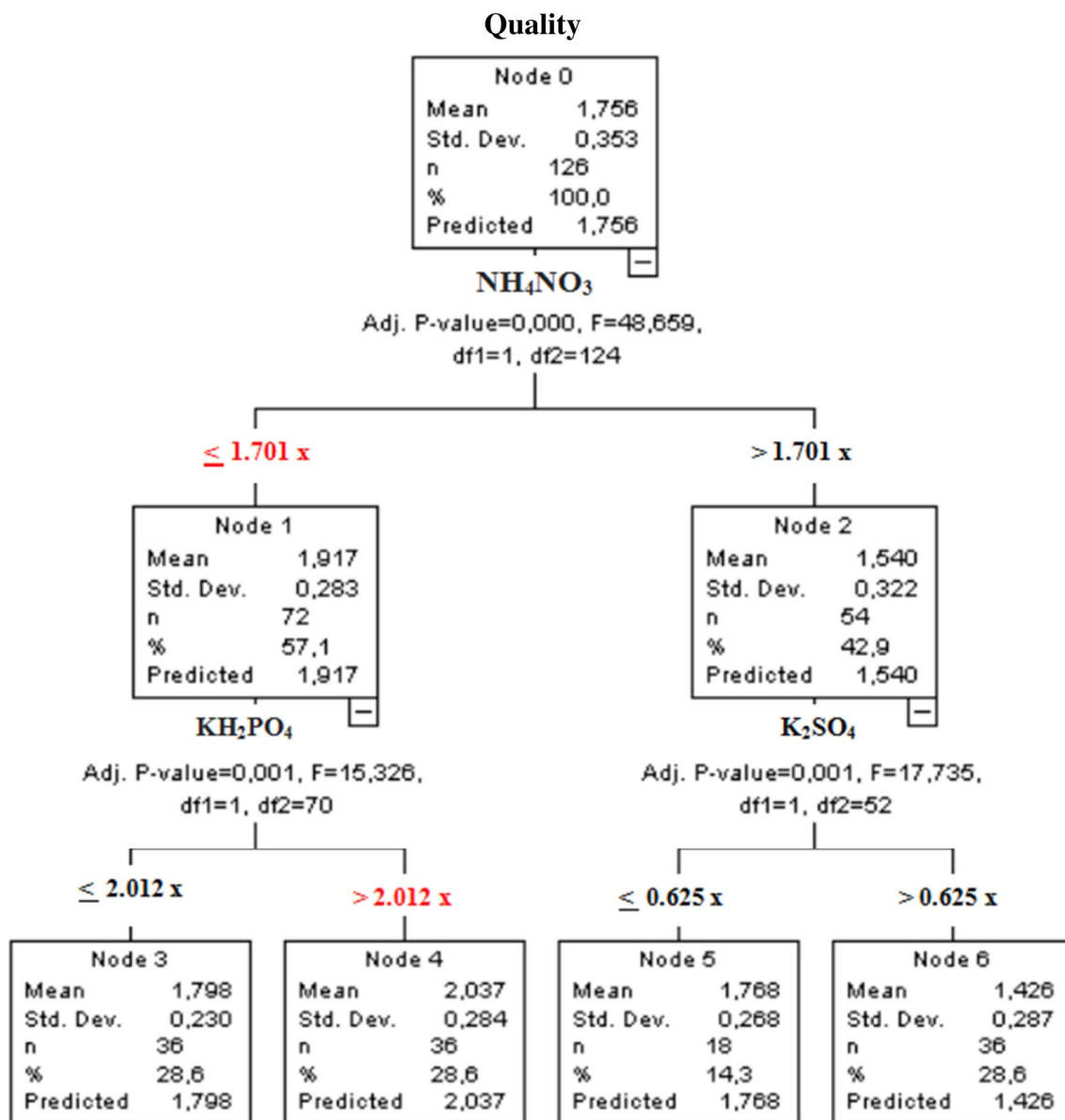
Quality

Fig. 3 The CHAID decision tree diagram for plant quality of 'Dorris', 'Wepster' and 'Zeta'. Nodes were determined by the significance of the factors. Salt cut-off values are × DKW

and 2) with respect to NH_4NO_3. Node 1 had significantly longer shoots (39.34 mm) than Node 2 (25.98 mm) (adj. p < 0.0001). Genotype affected shoot length within both Nodes 1 and 2 (adj. p < 0.001). Node 1 (NH_4NO_3 ≤1.701×) branched into two nodes by genotype. Node 3 ('Dorris' and 'Wepster') was divided into three nodes (Nodes 7, 8 and 9) based on KH_2PO_4. The greatest shoot length for these genotypes (39.44 mm) was with NH_4NO_3 ≤1.701× and KH_2PO_4 >2.75× (Node 9). Node 4 ('Zeta') on treatments with NH_4NO_3 ≤1.701× had an average shoot length of 50.79 mm, and was divided into three new nodes (Nodes 10, 11 and 12), based on $Ca(NO_3)_2$. The longest shoots (60.389 mm) for 'Zeta' were recorded from Node 12 with

NH_4NO_3 ≤1.701× and $CaCl_2$ >1.8×. A common salt concentration for the tested genotypes with ideal shoot length of 40 mm is suggested to be: NH_4NO_3 ≤1.701×, KH_2PO_4 >2.75×, $CaCl_2$ <1.738×.

Shoot length results varied between the two analyses. The CHAID data mining algorithm determined that KH_2PO_4 >2.75× and NH_4NO_3 ≤1.701× were critical cut-off values for the longest shoots of 'Dorris' and 'Wepster'. This was similar to RSM results. However, for 'Zeta' RSM indicated only NH_4NO_3 was important (Fig. 4; Table 2), while CHAID found $CaCl_2$ also to be significant (Fig. 5). Nas and Read (2004) defined a micropropagation medium for hybrid hazelnuts (*C. avellana* × *C. americana*) based

Fig. 4 Response surface graphs of mineral nutrient effects on shoot length (mm) of **a** 'Dorris', **b** 'Wepster' and **c** 'Zeta'. The shoot lengths (mm) were color coordinated from longest (*red-yellow*) to shortest (*green–blue*). The *red dot* represents the control with average shoot length of 50 mm. (Color figure online)
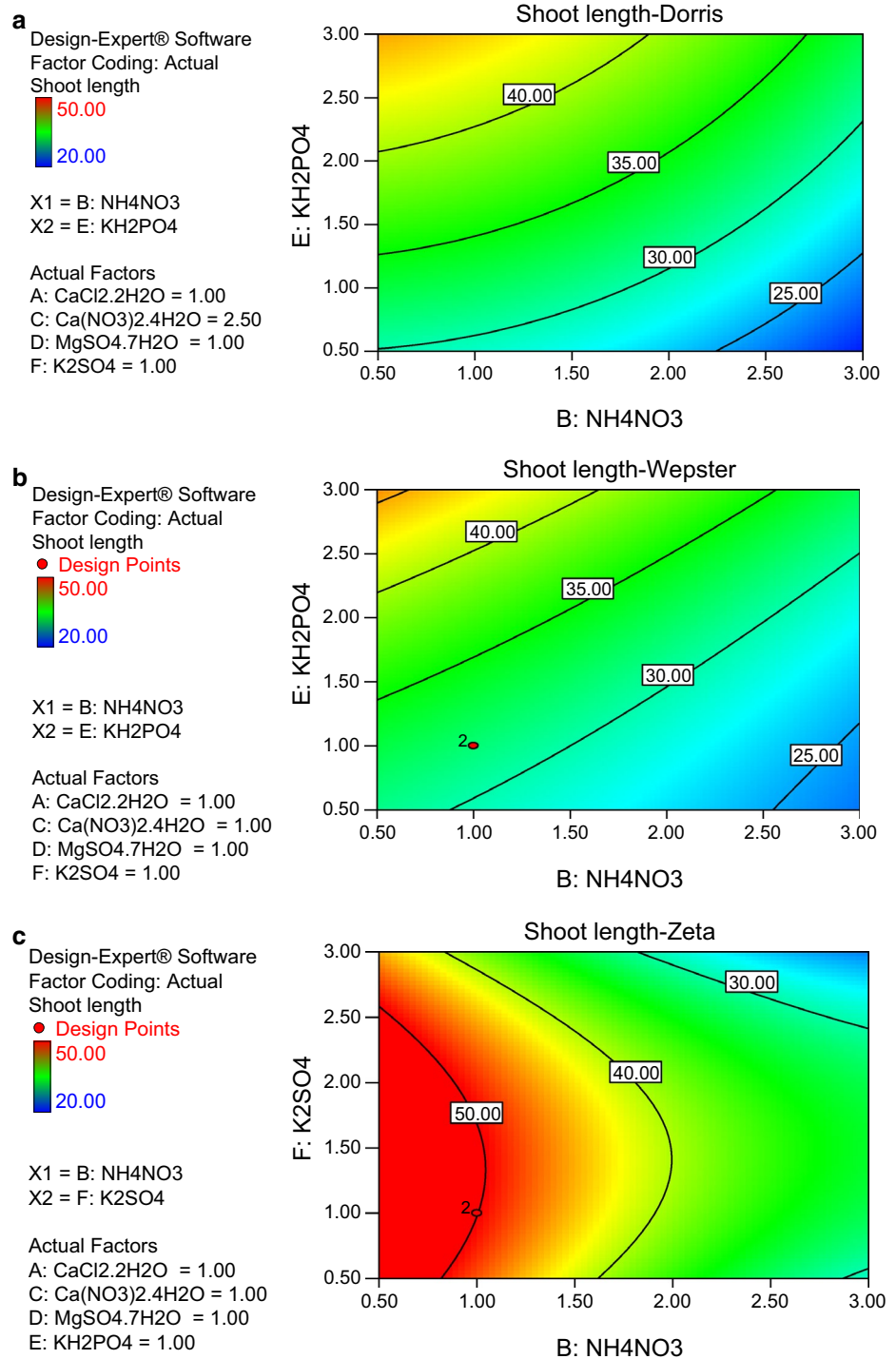
on the chemical composition of the hazelnut kernel. Better shoot elongation was observed on lower NH_4NO_3, $Ca(NO_3)_2$ and $CaCl_2$, as well as high $MgSO_4$ and KH_2PO_4 compared to DKW. The differences in salt requirements could be atrributed mainly to genotype (*C. avellana* versus hybrid hazelnuts), experimental design and the various statistical methods applied. Hand et al. (2014) found shoot length quite variable by genotype, however increased nitrogen factors and mesos were commonly involved.

Shoot number

RSM models for shoot number were significant ($p < 0.01$) for all of the hazelnut cultivars (Table 2). NH_4NO_3 was significant for all of the genotypes ($p < 0.01$), and low concentrations were required for high shoot number (Fig. 6). 'Dorris' produced high shoot multiplication (4.5) at the highest KH_2PO_4 and the lowest $MgSO_4$ amounts ($p < 0.05$). Low amounts of K_2SO_4 and very low $Ca(NO_3)_2$ increased shoot
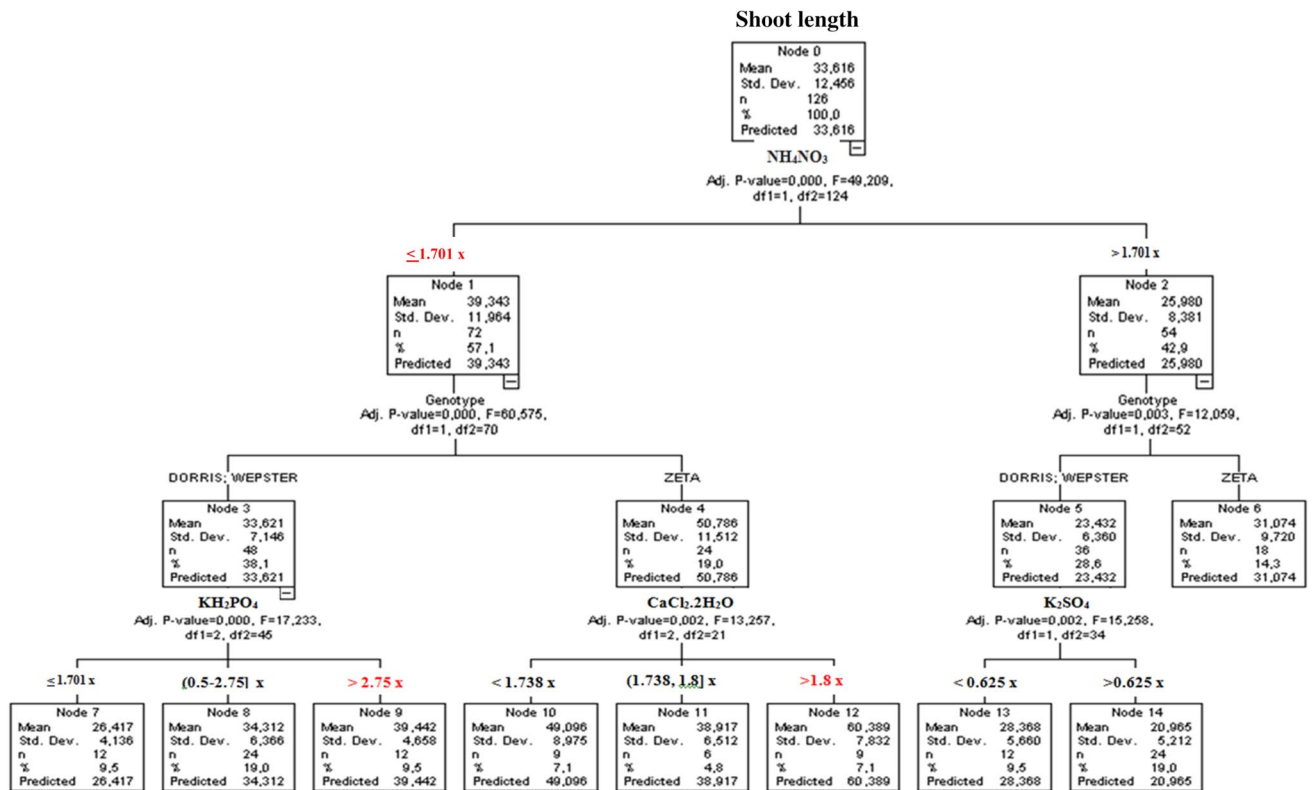
Shoot length

Fig. 5 The CHAID decision tree diagram for shoot length of 'Dorris', 'Wepster' and 'Zeta' hazelnuts. Nodes were determined by the significance of the factors. Salt cut-off values are × DKW

numbers for 'Dorris', depending on the amounts of NH_4NO_3 and $MgSO_4$ ($p < 0.05$) (Fig. 6a). A negative interaction of $NH_4NO_3 \times CaCl_2$ resulted in the highest shoot number (2.6) for 'Wepster' ($p < 0.001$), with the lowest NH_4NO_3 and the highest $CaCl_2$ and a low amount of K_2SO_4 (Fig. 6b). 'Zeta' showed very low $Ca(NO_3)_2$ and NH_4NO_3 requirements for high shoot number (2.4) ($p < 0.0001$) (Fig. 6c). Low $Ca(NO_3)_2$ and NH_4NO_3 were both important for increased shoot production in other *C. avellana* genotypes (Hand et al. 2014). Nas and Read (2004) suggested lower NH_4NO_3, $Ca(NO_3)_2$ and $CaCl_2$, as well as higher $MgSO_4$ and KH_2PO_4 for better multiplication compared to DKW.

The CHAID algorithm indicated that the most significant factor for shoot number was NH_4NO_3 (adj. $p < 0.001$), followed by $Ca(NO_3)_2$ (adj. $p < 0.01$) (Fig. 7). Genotype was an insignificant source of variation for multiplication. The Pearson correlation coefficient for shoot number was 0.653 ($p < 0.01$). All the shoots (Node 0) were split into three nodes (Nodes 1, 2 and 3) based on the response to NH_4NO_3 concentration. Node 1 ($NH_4NO_3 \leq 0.5\times$) shoots were affected by the $Ca(NO_3)_2$ concentration (average shoot number 2.197). Node 2 was those grown within the range of $0.5\times < NH_4NO_3 \leq 2.6\times$ (average shoot number 1.773), were also influenced by $Ca(NO_3)_2$ and KH_2PO_4 (adj. $p < 0.01$). Shoot numbers of plants exposed to $NH_4NO_3 > 2.6\times$ were

not affected by any other basal salt factor or genotype (adj. $p < 0.001$). The highest multiplication (2.46) was obtained in Node 4 with $NH_4NO_3 \leq 0.5\times$ and $Ca(NO_3)_2 \leq 1.725\times$ (Fig. 7).

Significant salt requirements for good multiplication varied depending on genotype in RSM (Fig. 6), whereas clear cut-off values of $NH_4NO_3 \leq 0.5\times$ and $Ca(NO_3)_2 \leq 1.725\times$ were suggested by the CHAID for best multiplication for the all tested genotypes (Fig. 7). Hand et al. (2014) using RSM suggested low amounts of NH_4NO_3 and $Ca(NO_3)_2$ for a higher multiplication rate of five *Corylus avellana* cultivars. The current study RSM results showed higher shoot number (4.5 versus 2.5) for 'Dorris' than Hand et al. (2014), which could be attributed to the higher concentration range of KH_2PO_4 within the experimental design ($3\times$ versus $1.5\times$), as well as to the differences in micro nutrient concentrations used in both studies.

Callus

Responses of 'Wepster' and 'Zeta' were significant for callus formation in RSM ($p < 0.05$), and the most influential factor was NH_4NO_3 ($p < 0.001$) (Table 2). KH_2PO_4, K_2SO_4, NH_4NO_3 and $MgSO_4$ were considered significant for callus of 'Zeta' ($p < 0.05$). An interaction of $NH_4NO_3 \times K_2SO_4$

Fig. 6 Response surface graphs of mineral nutrient effects on shoot number of **a** 'Dorris', **b** 'Wepster' and **c** 'Zeta'. The shoot numbers were color coordinated from most (*red-yellow*) to fewer shoots (*green–blue*). The *red dot* represents the control with 2.2 average shoot number. (Color figure online)
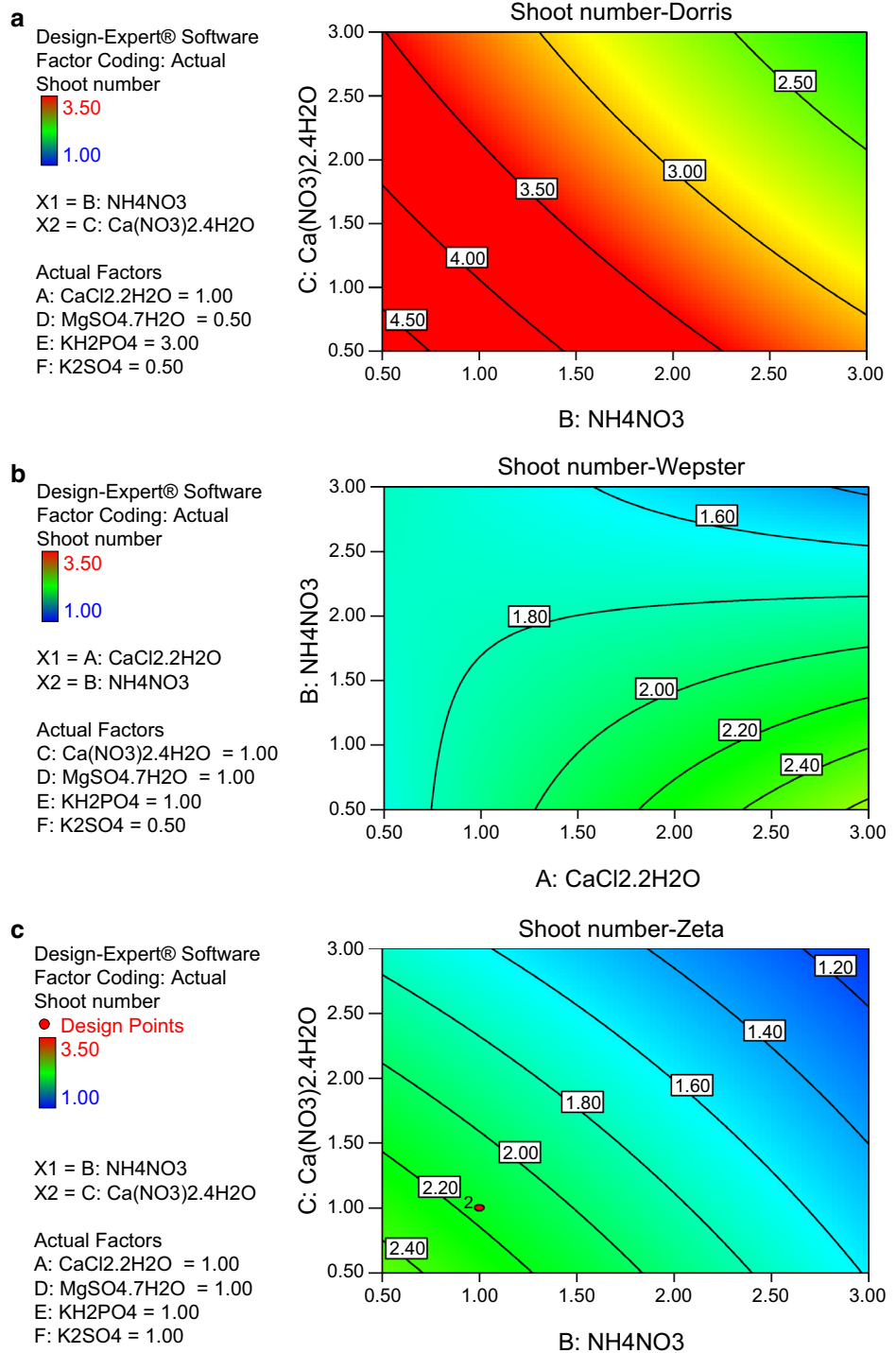
affected callus formation of 'Wepster' ($p < 0.05$) (Table 2). The highest concentrations of NH_4NO_3 and K_2SO_4 resulted in the least callus (rating of 2.6) for 'Wepster' (Supplement 1). 'Zeta' required the highest amounts of NH_4NO_3, K_2SO_4 and the lowest concentrations of KH_2PO_4 and $MgSO_4$ for low callus (rating of 2.8) (Supplement 1b). These requirements contradict those needed for good plant quality (Fig. 1).

Mineral nutrient factors NH_4NO_3, $MgSO_4$, K_2SO_4 along with genotype were statistically defined as affecting callus

formation using a CHAID decision tree (Supplement 2). The Pearson correlation coefficient was strongly significant at 0.70. At the top of the decision tree diagram, the root node showed an average callus rating of 2.18 for all the analyzed shoots. NH_4NO_3 had a dominant effect on callus formation for all of the hazelnut genotypes (adj. $p < 0.001$) and the shoots were classified into three Nodes (Supplement 2). Node 1 was the group of hazelnut shoots exposed to $NH_4NO_3 \leq 0.5\times$ (mean rating 1.986); Node 2 within the
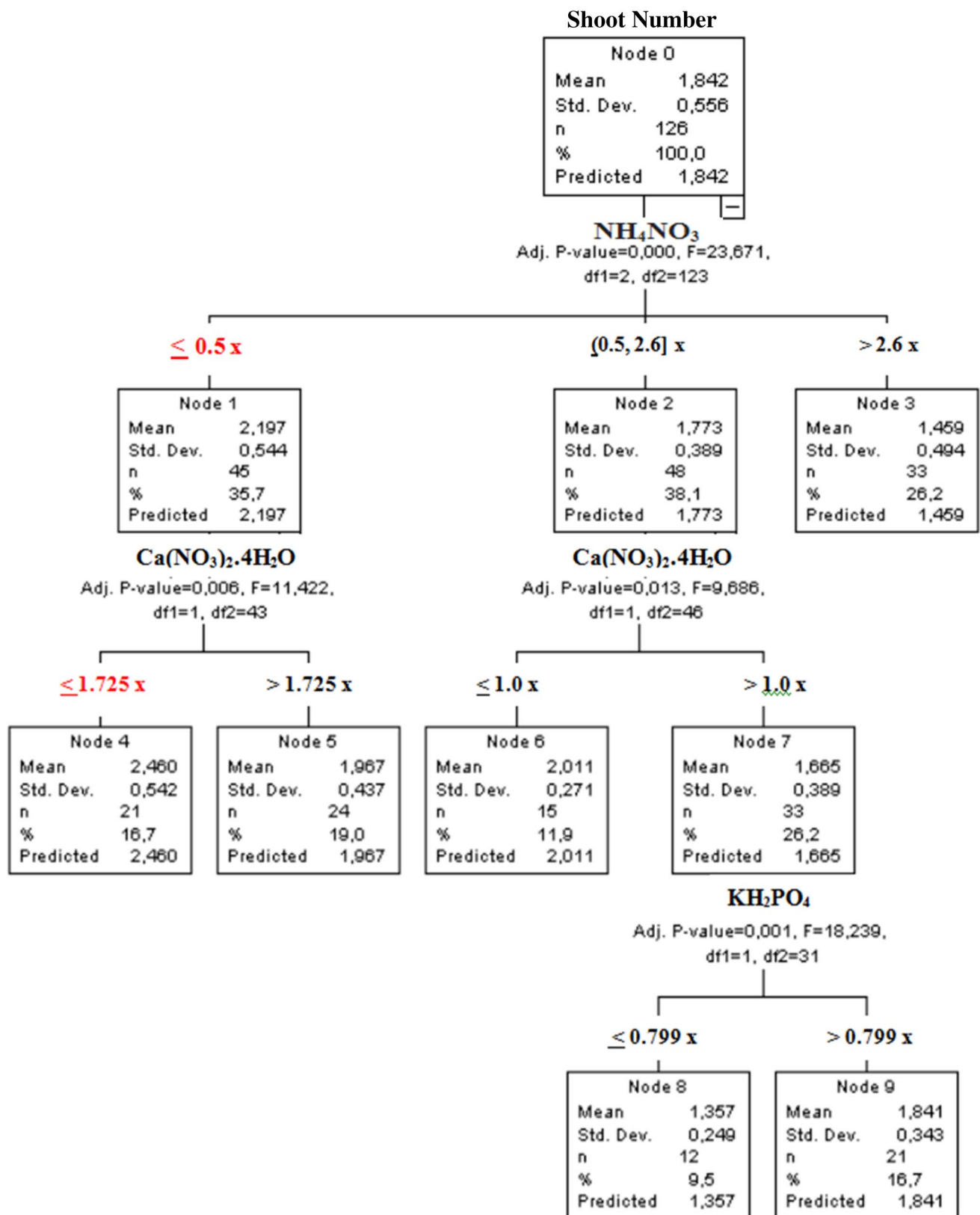
Shoot Number

Node 0
Mean	1,842
Std. Dev.	0,556
n	126
%	100,0
Predicted	1,842

NH_4NO_3
Adj. P-value=0,000, F=23,671,
df1=2, df2=123

≤ 0.5 x   [0.5, 2.6] x   > 2.6 x

Node 1
Mean	2,197
Std. Dev.	0,544
n	45
%	35,7
Predicted	2,197

Node 2
Mean	1,773
Std. Dev.	0,389
n	48
%	38,1
Predicted	1,773

Node 3
Mean	1,459
Std. Dev.	0,494
n	33
%	26,2
Predicted	1,459

$Ca(NO_3)_2.4H_2O$
Adj. P-value=0,006, F=11,422,
df1=1, df2=43

$Ca(NO_3)_2.4H_2O$
Adj. P-value=0,013, F=9,686,
df1=1, df2=46

≤ 1.725 x   > 1.725 x   ≤ 1.0 x   > 1.0 x

Node 4
Mean	2,460
Std. Dev.	0,542
n	21
%	16,7
Predicted	2,460

Node 5
Mean	1,967
Std. Dev.	0,437
n	24
%	19,0
Predicted	1,967

Node 6
Mean	2,011
Std. Dev.	0,271
n	15
%	11,9
Predicted	2,011

Node 7
Mean	1,665
Std. Dev.	0,389
n	33
%	26,2
Predicted	1,665

KH_2PO_4
Adj. P-value=0,001, F=18,239,
df1=1, df2=31

≤ 0.799 x   > 0.799 x

Node 8
Mean	1,357
Std. Dev.	0,249
n	12
%	9,5
Predicted	1,357

Node 9
Mean	1,841
Std. Dev.	0,343
n	21
%	16,7
Predicted	1,841

Fig. 7 The CHAID decision tree diagram for shoot number of 'Dorris', 'Wepster' and 'Zeta'. Nodes were determined by the significance of the factors. Salt cut-off values are × DKW

range of $0.5\times < NH_4NO_3 \leq 1.701\times$ (mean rating 2.136), and Node 3 within $NH_4NO_3 > 1.701\times$ (mean rating 2.364). Callus formation of shoots within Node 2 ($0.500\times < NH_4NO_3 \leq 1.701\times$) was not affected by any other basal salt factor or genotype. Genotype was an important factor for callus formation in Node 3 ($NH_4NO_3 > 1.701\times$) (adj. $p < 0.01$), which branched into Nodes 7 and 8. Callus of 'Dorris' exposed to $NH_4NO_3 > 1.701\times$ (Node 7) was not affected by any other salt factor. The cut-off values for callus of 'Wepster' and 'Zeta' hazelnuts were estimated as $1.701\times NH_4NO_3$ and $0.5\times K_2SO_4$. Callus production of 'Wepster' and 'Zeta' within group $NH_4NO_3 > 1.701\times$ (Node 8), was affected by K_2SO_4 with a mean rating of 2.482 (adj. $p < 0.05$). The least callus (2.585) was obtained in Node 10, which was the group of 'Wepster' and 'Zeta' exposed to $NH_4NO_3 > 1.701\times$ and $K_2SO_4 > 0.5\times$. In addition, Node 9 ('Wepster' and 'Zeta' in the range of $NH_4NO_3 > 1.701\times$ and $K_2SO_4 \leq 0.5\times$) produced the mean callus rating of 2.21. Despite resulting in less callus, both of those nodes were poor for overall shoot quality. Terminal node 5 (average callus 2.2) which is within the range of $NH_4NO_3 \leq 0.5\times$ and $0.95\times < MgSO_4 \leq 1.762\times$, is consistent with the requirements for good plant quality.

For callus, RSM was generally consistent with the results of CHAID, but didn't provide a significant model for 'Dorris' (Table 2). Hand et al. (2014) suggested high amounts of NH_4NO_3 and $Ca(NO_3)_2$ for less callus formation of 'Dorris'. In the current study, very high concentrations of K_2SO_4 and very low amounts of KH_2PO_4 and $MgSO_4$ were required for less callus formation of 'Zeta' (Supplement 1a). These differences could be attributed to differences in the statistical background of RSM and CHAID methodologies. All the cultivars were simultaneously analyzed using CHAID algorithm, which provided special information about the classification of plants giving similar or different responses to the combinations defined by RSM design.

In the RSM analysis, great variation was apparent between hazelnut cultivars in terms of quality, shoot length, shoot number and callus (Hand et al. 2014; Hand and Reed 2014) where cultivar was not included as a factor, but the CHAID analysis revealed the similarities and differences between the cultivars. In earlier RSM based studies, cultivar variation made it extremely difficult to formulate a common hazelnut micropropagation medium (Hand et al. 2014; Hand and Reed 2014). The results of the tree-based CHAID algorithm (Figs. 3, 5, 7; Supplement 2) in the current study closely mirrored the graphical results of RSM (Figs. 1, 4, 6; Supplement 1), but were easier to interpret. Alteration of the concentration ranges of nutrient factors within the RSM design might also improve the optimization process. Previous studies showed that a neuro-fuzzy logic data mining algorithm was useful for characterizing predictive nutrient factors that directly correlated with plant responses to tissue culture medium (Alanagh et al. 2014; Gago et al. 2011),

but the exact cut-off amounts of nutrients, and how their significance on the response could change based on the factor concentration range, were not determined. Nas and Read (2004) hypothesized that the seed mineral and organic composition could be ideal starting point for tissue culture medium optimization. However, the seed nitrogen content was found toxic for the in-vitro hazelnuts.

Overall shoot quality, shoot length and multiplication improved with lower NH_4NO_3 concentrations. In the CHAID analysis low NH_4NO_3 provided good shoot quality, shoot length, multiplication and reduced callus (Figs. 3, 5, 7; Supplement 2). In RSM it was more complicated to determine a common formula for high quality, shoot length and multiplication versus a low callus model (Fig. 1). However, in general, analysis using RSM indicated that increased KH_2PO_4 and lower K_2SO_4 and NH_4NO_3 concentrations improved the quality of all three cultivars (Figs. 1, 2). We can more easily conclude from the prediction trees of the CHAID algorithm (Figs. 3, 5, 7; Supplement 2) that $NH_4NO_3 \leq 0.5\times$, $KH_2PO_4 > 2.75\times$, $Ca(NO_3)_2 < 1.725\times$ are the required critical cut-offs for optimum growth medium of the three hazelnut genotypes evaluated. Based on this analysis the new medium amounts were set at: $0.5\times NH_4NO_3$, $3\times KH_2PO_4$, $1.5\times Ca(NO_3)_2$. The other salt factors that were analyzed could be set at the standard DKW concentrations ($1\times$) and remain within the indicated cut-off ranges.

Conclusions

Computer generated optimal design, like RSM, is an excellent tool for reducing treatment numbers compared to traditional factorial designs, and analyzing the resulting data with CHAID provides clear direction for developing a single optimal medium or a small number of suitable media for a range of genotypes. These advanced data mining approaches could be implemented to deduce optimum cut off-values of nutrient factors from mineral nutrition studies. CHAID is a novel and promising approach in tissue culture medium optimization. It provides a visual tree and exact cut-offs of the significant nutrients which makes it easier to define optimal concentrations of the nutrient salts. Evaluating in vitro culture data with the CHAID algorithm would provide clearer evaluation of the responses of the in vitro plants to the tested variables. The property of CHAID allowing the analysis of all genotypes together would contribute to developing one or a few optimal media for many genotypes rather than many cultivar-specific formulations.

MA in the Department of Horticulture, Oregon State University, Corvallis, OR.

References

Alanagh EN, Garoosi G-a, Haddad R, Maleki S, Landín M, Gallego PP (2014) Design of tissue culture media for efficient Prunus rootstock micropropagation using artificial intelligence models. Plant Cell Tissue Organ Cult 117(3):349–359

Ali M, Eyduran E, Tariq MM, Tirink C, Abbas F, Bajwa MA, Baloch MH, Nizamani AH, Waheed A, Awan MA, Shah SH, Ahmad Z, Jan S (2015) Comparison of Artificial neural network and decision tree algorithms used for predicting live weight at post weaning period from some biometrical characteristics in harnai sheep. Pak J Zool 47(6):1579–1585

Anderson MJ, Whitcomb PJ (2005) RSM simplified: optimizing processes using response surface methods for design of experiments. Productivity Press, New York

Bradley N (2007) The Response Surface Methodology. Dissertation: Indiana University South Bend

Compton M, Mize C (1999) Statistical considerations for in vitro research: I—birth of an idea to collecting data. In Vitro Cell Dev Biol Plant 35(2):115–121

Design-Expert (2010) Stat-Ease, Inc., Minneapolis

Driver JA, Kuniyuki AH (1984) In vitro propagation of Paradox walnut rootstock. HortScience 19(4):507–509

Gago J, Perez-Tornero O, Landin M, Burgos L, Gallego PP (2011) Improving knowledge of plant tissue culture and media formulation by neurofuzzy logic: a practical case of data mining using apricot databases. J Plant Physiol 168(15):1858–1865

Hand C, Reed BM (2014) Minor nutrients are critical for the improved growth of Corylus avellana shoot cultures. Plant Cell Tissue Organ Cult 119(2):427–439

Hand C, Maki S, Reed B (2014) Modeling optimal mineral nutrition for hazelnut micropropagation. Plant Cell Tissue Organ Cult 119(2):411–425

Hébert M, Collin-Vézina D, Daigneault I, Parent N, Tremblay C (2006) Factors linked to outcomes in sexually abused girls: a regression tree analysis. Compr Psychiatry 47:443–455

Ibañez MA, Martin C, Pérez C (2003) Alternative statistical analyses for micropropagation: a practical case of proliferation and rooting phases in Viburnum opulus. In Vitro Cell Dev Biol Plant 39(5):429–436

Mize C, Koehler K, Compton M (1999) Statistical considerations for in vitro research: II —data to presentation. In Vitro Cell Dev Biol Plant 35(2):122–126

Montgomery DC (2005) Design and analysis of experiments: response surface method and designs. Wiley, New Jersey

Nas MN, Read PE (2004) A hypothesis for the development of a defined tissue culture medium of higher plants and micropropagation of hazelnuts. Sci Hortic 101(1–2):189–200

Nas MN, Eskridge K, Read P (2005) Experimental designs suitable for testing many factors with limited number of explants in tissue culture. Plant Cell Tissue Organ Cult 81(2):213–220

Niedz RP, Evens TJ (2007) Regulating plant tissue growth by mineral nutrition. In Vitro Cell Dev Biol Plant 43(4):370–381

Niedz RP, Hyndman SE, Evens TJ (2007) Using a gestalt to measure the quality of in vitro responses. Sci Hortic 112(3):349–359

Rashidi S, Ranjitkar P, Hadas Y (2014) Modeling bus dwell time with decision tree-based methods. Transport Res Rec 2418:74–83

Reed BM, Wada S, DeNoma J, Niedz RP (2013) Improving in vitro mineral nutrition for diverse pear germplasm. In Vitro Cell Dev Biol Plant 49:343–355

SPSS (2013) Statistics for Windows, Version 22.0. IBM Corp, Armonk

Statistics-Solutions (2016) CHAID. http://www.statisticssolutions.com/non-parametric-analysis-chaid/